

nature



THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

Ten
people
who
mattered
this year
PAGE 459

ONE YEAR. TEN STORIES.

GENE EDITING

CRAZY FOR CRISPR

Takes from a pioneer
of the technology

PAGE 469

SILICON PHOTONICS

CHIPS WITH EVERYTHING

A 'system-on-a-chip'
microprocessor

PAGES 482 & 534

CALENDAR

EVENTS DIRECTORY 2016

The Nature guide to global
science events and courses

BACK PAGES & NATURE.COM

NATUREASIA.COM

24/21 December 2015

VOL 528 NO 7581

THIS WEEK

EDITORIALS

INTERFERENCE On fairy lights, broadband and global migration **p.436**

WORLD VIEW What you didn't hear about the new climate deal **p.437**



TRUNK CALLING Daughters follow in female elephants' tracks **p.438**

Fishy limits

The European Union has set a worrying trend by ignoring scientific advice on overfishing. It must put long-term sustainability plans ahead of short-term political gains.

Fish have a memory capacity that goes far beyond what they are usually given credit for, but do European politicians? If not, the Ghost of Christmas Past could remind ministers of any number of grim scenes from recent years: the decades of overfishing, the large decline in stocks such as cod, and the dire and repeated warnings from scientists that ocean resources are being depleted faster than they can recover.

With a little seasonal flexibility, the Ghost could even show politicians the agreement they signed in 2013 to use proper scientific advice when setting annual fishing quotas, formally known as total allowable catches (TACs). And, if they are still refusing to wake up, the Ghost could take them on a brief trip back to last week, when the policymakers turned their back on that promise.

Never mind the Ghost of Christmas Present: a meeting last week in Brussels saw the giving and receiving of Christmas presents from the politicians to each other, to their domestic fishing industries and to vocal lobby groups. Although the headline news celebrated the recovery of some iconic fish stocks — North Sea cod among them — and the increased licence that fishermen again have to scoop them up in greater numbers, the story beneath the surface was not so happy. For many species, scientific advice was again ignored, and TACs that look unsustainable were agreed.

Cod in the Kattegat Sea, the shallow and treacherous waters between Denmark and Sweden, are still struggling, and face a much more uncertain future than their cousins in the North Sea. The meeting last week offered them little cheer. The agreed TAC is some three times the size of the quota recommended by the International Council for the Exploration of the Sea, the scientific body that advises the European Union. Celtic Sea cod and Southern hake are among the other fish for which scientists had proposed stricter limits than the politicians agreed, and which are now left exposed to overfishing.

One reason why the outcome of the Brussels meeting is so disappointing is that it comes after encouraging signs that the message on overfishing was finally getting through.

Research published last month shows that since 2001, European fisheries TACs have been an average of 20% higher than scientific advice suggested (G. Carpenter *et al. Mar. Policy* **64**, 9–15; 2016). But the picture is improving. The same study found that whereas fishing was 33% above the recommended level in 2001, it was only 7% higher in 2015. There is more scrutiny on fisheries, more public interest and seemingly more political will to tackle the problem than there has been in the past. When promising to respect the scientific advice on quotas in 2013, Europe also pledged to move towards catches based on a different, more ecological, measure of stock health called maximum sustainable yield by 2020.

The message sent last week by the willingness of the European policymakers to ignore scientific advice places a question mark over whether progress can be sustained, and the 2020 target reached.

Despite the recovery of some landmark species (only after, it should be said, draconian and last-ditch fishing curbs were placed on them), study after study has shown that many European fish species remain in peril. Just last week, the Marine Stewardship Council, a non-profit organization dedicated to tackling overfishing, suspended all five cod fisheries in the Eastern Baltic Sea from its scheme that awards sustainable status to fish products.

“Sustainable fishing offers more security than haphazard political agreements.”

Fishing is a difficult political problem. One analysis has found that overfishing is more likely where fish stocks are large and exploited by a number of different countries (see go.nature.com/mhx6q4).

Low quotas have a genuine social and economic impact on a vulnerable sector and the people who work in it. It is natural that politicians want to protect jobs and maintain livelihoods. But scientists and conservationists want that too. They just think a little further ahead. Ultimately, sustainable fishing offers more security than haphazard political agreements made behind closed doors from year to year.

Announcing the most recent round of TACs, Karmenu Vella, the EU fisheries commissioner, said: “We cannot jeopardise the longer term sustainability for the shorter term considerations.” No one could disagree with that. Vella added: “We are on track in our sustainability targets.” Universal agreement for that statement will be harder to find. The Ghost of Christmas Yet to Come awaits. ■

Quantum leap

Physicists can better study the quantum behaviour of objects on the atomic scale.

Erwin Schrödinger was an interesting man. Not only did he conceive a most imaginative way to (theoretically) kill a cat, he was in a constant state of superposition between monogamy and not. He shared a household with one wife and one mistress. (Although he got into trouble at Oxford for this unconventional lifestyle, it didn't pose a problem in largely Catholic Dublin.) Just like the chemist Albert Hofmann, who tried LSD (lysergic acid diethylamide) on himself first, Schrödinger might have pondered how it would feel for a person to be in a genuine state of quantum superposition. Or even how a cat might feel.

In principle, quantum mechanics would certainly allow for Schrödinger, or any of us, to enter a state of quantum superposition.

That is, according to quantum theory, a large object could be in two quantum states at the same time. It is not just for subatomic particles.

Everyday experience, of course, indicates that big objects behave classically. In special labs and with a lot of effort, we can observe the quantum properties of photons or electrons. But even the best labs and greatest efforts are yet to find them in anything approaching the size of a cat.

Could they be found? The question is more than head-in-the-clouds philosophy. One of the most important experimental questions in quantum physics is whether or not there is a point or boundary at which the quantum world ends and the classical world begins.

A straightforward approach to clarifying this question is to experimentally verify the quantum properties of ever-larger macroscopic objects. Scientists find these properties in subatomic particles when they confirm that the particles sometimes behave as a wave, with characteristic peaks and dips. Likewise, lab set-ups based on the principle of quantum interference, using many mirrors, lasers and lenses, have successfully found wave behaviour in macromolecules that are more than 800 atoms in size.

Other techniques could go larger. Called atom interferometers, they probe atomic matter waves in the way that conventional interferometers measure light waves. Specifically, they divide the atomic matter wave into two separate wave packets, and recombine them at the end. The sensitivity of these devices is related to how far apart they can perform this spatial separation. Until now, the best atomic interferometers could put the wave packets about 1 centimetre apart.

On page 530 of this issue, physicists demonstrate an astonishing advance in this regard. They show quantum interference of atomic wave packets that are separated by 54 centimetres. Although this does not mean that we have an actual cat in a state of quantum superposition, at least a cat could now comfortably take a nap between the two

branches of a superposed quantum state. (No cats were harmed in the course of these experiments.)

Making huge molecules parade their wave nature and constructing atom interferometers that can separate wave packets by half a metre are extraordinary experimental achievements. And the technology coming from these experiments has many practical implications: atom interferometers splendidly measure acceleration, which means that

they could find uses in navigation. And they would make excellent detectors for gravitational waves, because they are not sensitive to seismic noise.

“A cat could now take a nap between the two branches of a superposed quantum state.”

Schrödinger was more of a philosopher than an engineer, so it is plausible that he would not have taken that much interest in the practical ramifications of his theory.

However, he would surely have clapped his hands at the prospect that experimenters could one day induce large objects to have quantum properties. And there are plenty of proposals for how to ramp up the size of objects with proven quantum behaviour: a microscopic mirror in a quantum superposition, created through interaction with a photon, would involve about 10^{14} atoms. And, letting their imaginations run wild, researchers have proposed a method to do the same with small biological structures such as viruses.

To be clear, science is not close to putting a person or a cat into quantum superposition. Many say that, because of the way large objects interact with the environment, we will never be able to measure a person's quantum behaviour. But it's Christmas, so indulge us. If we could, and if we could be aware of such a superposition state, then how would we feel? Because 'feeling' would amount to measuring the wave function of the object, and because measuring causes the wave function to collapse, it should really feel like, well, nothing — or perhaps just a grin. ■

Light relief

Nature digs into the rumours about the effect of festive illuminations on wireless fidelity.

At the end of the year, it is natural to reflect on the many science success stories of 2015. There was the forging of a climate-change agreement in Paris, and the incredible pictures of Pluto beamed back by the New Horizons spacecraft (for more, see our end-of-year review starting on page 448). Beware, though, for the road of progress is bumpy, and new and old technology can clash.

Christmas can break the Internet, the UK newspapers nearly reported this month. Researchers have found that twinkling fairy lights on a household Christmas tree can interfere with the wireless signal between a router and internet-connected devices.

In Britain, the telephony and airwaves regulator Ofcom released a smartphone app so that people can assess just how bad this seasonal effect is. We at *Nature* know what's expected of us, so we downloaded the app and put it through its paces.

First, the control test. The Nature Towers Wi-Fi was just fine before we illuminated the office Christmas tree, and — to the relief of all — remained completely unaffected once the halls were decked with the requisite tinsel, mistletoe, boughs of holly and festive lighting. Still, before you eat another mince pie and check the online weather forecast for snow, know that the Wi-Fi was seriously compromised by unknown forces once the illuminations had been switched off for the night. What could have been going on?

As Andrew Smith writes on *The Conversation*, your festive illuminations might indeed interfere with your Wi-Fi, but they would have to be very powerful — much more so than other household features such as

microwaves or fluorescent lights (see go.nature.com/fqy5mr).

The *Daily Mail* newspaper can always be relied on for inventive scientific answers and did not disappoint. Perhaps, it says, goldfish are sabotaging the Wi-Fi? Water, it points out, absorbs radio waves, so you shouldn't place a router near a fish tank, nor (we suppose) in one.

The story, although little more than a sprinkling of seasonal fluff on the tail end of the year in science, does illustrate more serious matters — the many factors, perhaps small and even undetectable, that can throw an experiment.

We all know colleagues whose Southern blots come out like Rorschach tests and who have to rely on the one lab technician who has 'the touch'. *Nature* argues strongly for reproducibility and that experimental details, no matter how small, should be set out for all to see. We have launched a string of publications and platforms to help researchers to do this: *Nature Methods*, *Nature Protocols*, *Scientific Data* and *Protocol Exchange*. However, when one is working just beyond the cutting edge, other factors might be at play — on the edge of detectability and beyond. One of last year's highlights was the discovery, after years of careful testing, that migrating birds can be disoriented by the electromagnetic 'smog' produced by human activity (S. Engels *et al.* *Nature* **509**, 353–356; 2014).

This finding sits in a contentious field in which researchers seek to explain the seemingly impossible feat in which animals detect and transduce the very weak signals generated by Earth's magnetic field. Festive bulbs are a mere drop in the electromagnetic ocean, from the devices around us to the photons that bring messages from the edge of the cosmos.

In the time it has taken you to read this, about 600 trillion neutrinos will have passed through your body, as well as uncounted dark-matter particles, and perhaps even some schleptons, snoozons, axions and other particles of which science has as no knowledge, yet. That is what next year is for. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunqv

CORRECTION

The Editorial 'Fishy limits' (*Nature* **528**, 435; 2015) wrongly implied that the European Commission had set the fishing quotas. They were set by the Council of Ministers.



Talks in the city of light generate more heat

Rather than relying on far-off negative-emissions technologies, Paris needed to deliver a low-carbon road map for today, argues Kevin Anderson.

The climate agreement delivered earlier this month in Paris is a genuine triumph of international diplomacy. It is a tribute to how France was able to bring a fractious world together. And it is testament to how assiduous and painstaking science can defeat the unremitting programme of misinformation that is perpetuated by powerful vested interests. It is the twenty-first century's equivalent to the victory of heliocentrism over the inquisition. Yet it risks being total fantasy.

Let's be clear, the international community not only acknowledged the seriousness of climate change, it also demonstrated sufficient unanimity to define it quantitatively: to hold "the increase in ... temperature to well below 2°C ... and to pursue efforts to limit the temperature increase to 1.5°C".

To achieve such goals demands urgent and significant cuts in emissions. But rather than requiring that nations reduce emissions in the short-to-medium term, the Paris agreement instead rests on the assumption that the world will successfully suck the carbon pollution it produces back from the atmosphere in the longer term. A few years ago, these exotic Dr Strangelove options were discussed only as last-ditch contingencies. Now they are Plan A.

Governments, prompted by their advisers, have plumped for BECCS (biomass energy carbon capture and storage) as the most promising 'negative-emissions technology'.

What does BECCS entail? Apportioning huge swathes of the planet's landmass to the growing of bioenergy crops (from big trees to tall grasses) — which absorb carbon dioxide through photosynthesis as they grow. Periodically, these crops are harvested, processed for worldwide travel and shipped around the globe before finally being combusted in thermal power stations. The CO₂ is then stripped from the waste gases, compressed (almost to a liquid), pumped through large pipes over potentially very long distances and finally stored deep underground in various geological formations (from exhausted oil and gas reservoirs through to saline aquifers) for a millennium or so.

The unquestioned reliance on negative-emission technologies to deliver on the Paris goals is the greatest threat to the new agreement. Yet BECCS, or even negative-emission technologies, received no direct reference throughout the 32-page package. Despite this, the framing of the 2°C goal and, even more, the 1.5°C one, is premised on the massive uptake of BECCS some time in the latter half of the century. Disturbingly, this is also the case for most of the temperature estimates ascribed to the outcome of the voluntary emissions cuts made by nations before the Paris meeting.

The scale of the assumption is breathtaking. It would be the equivalent of decades of planting and harvesting of energy crops over an area of

one to three times that of India. At the same time, the aviation industry envisages powering its planes with biofuel, the shipping industry is seriously considering biomass to propel its ships and the chemical sector sees biomass as a potential feedstock — and by then there will be 9 billion or so human mouths to feed. This crucial assumption deserves wider scrutiny.

Relying on the promise of industrial-scale negative-emissions technologies to balance the carbon budget was not the only option available in Paris — at least in relation to 2°C.

Reducing emissions in line with 2°C remains a viable goal — just. But rather than rely on post-2050 BECCS, deciding to pursue this alternative approach would have begged profound political, economic and social questions. Questions that undermine a decade of mathematically nebulous green-growth and win-win rhetoric, and questions that the politicians have decided cannot be asked.

Move away from the cosy tenets of contemporary economics and a suite of alternative measures comes into focus. Technologies, behaviours and habits that feed energy demand are all amenable to significant and rapid change. Combine this with an understanding that just 10% of the population is responsible for 50% of emissions, and the rate and scope of what is possible becomes evident.

The allying of deep and early reductions in energy demand with rapid substitution of fossil fuels by zero-carbon alternatives frames a 2°C agenda that does not rely on negative emissions. So why was this real opportunity muscled out by the economic bouncers in Paris? No doubt there are many elaborate and nuanced explanations —

but the headline reason is simple. In true Orwellian style, the political and economic dogma that has come to pervade all facets of society must not be questioned. For many years, green-growth oratory has quashed any voice with the audacity to suggest that the carbon budgets associated with 2°C cannot be reconciled with the mantra of economic growth.

I was in Paris, and there was a real sense of unease among many scientists present. The almost euphoric atmosphere that accompanied the circulation of the various drafts could not be squared with their content. Desperate to maintain order, a club of senior figures and influential handlers briefed against those who dared to say so — just look at some of the Twitter discussions!

It is pantomime season and the world has just gambled its future on the appearance in a puff of smoke of a carbon-sucking fairy godmother. The Paris agreement is a road map to a better future? Oh no it's not. ■

Kevin Anderson is deputy director of the Tyndall Centre for Climate Change Research, UK.
e-mail: kevin.anderson@manchester.ac.uk. Twitter @KevinClimate

THE ALMOST
EUPHORIC
ATMOSPHERE THAT
ACCOMPANIED
THE DRAFTS
COULD NOT BE
SQUARED WITH THE
CONTENT.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/pmyruf

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

ATMOSPHERIC SCIENCE

High-energy flashes in the sky

High-frequency electromagnetic flashes, once thought to be rare, may go off regularly alongside lightning in the atmosphere.

In 1994, physicists discovered flickers of γ -rays associated with lightning storms, but had seen relatively few of them. A team led by Nikolai Østgaard at the University of Bergen, Norway, looked for more of these terrestrial γ -ray flashes in data taken by the Reuven Ramaty High Energy Solar Spectroscopic Imager satellite in 2006 and 2012.

The researchers found nearly 200 flashes; these may be more common, and release more energy into the atmosphere, than scientists had suspected.

Geophys. Res. Lett.

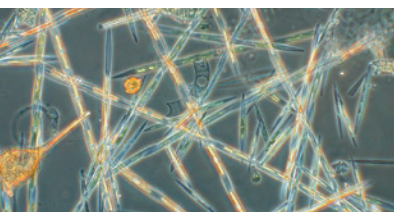
<http://doi.org/97g> (2015)

ECOTOXICOLOGY

Toxin clouds sea-lion memory

Toxic algal blooms could be impairing the memory and navigation of California sea lions (*Zalophus californianus*), possibly interfering with how they forage.

Domoic acid is a naturally occurring neurotoxin that is released by certain algae (such as *Pseudo-nitzschia* species, **pictured**). It is known to damage the hippocampus, a key memory centre in the brain, but its effects on



behaviour have been unclear. Between 2009 and 2011, Peter Cook, now at Emory University in Atlanta, Georgia, and his team studied 30 wild sea lions that were undergoing veterinary rehabilitation off the California coast. Using magnetic resonance imaging, the team found that animals with greater neurotoxin damage to the upper right portion of the hippocampus performed worse than animals with less-damaged brains on several spatial memory tasks, such as recalling the location of a bucket of fish.

Because these animals rely heavily on foraging, the toxin

could affect their survival in the wild, the authors say.

Science 350, 1545–1547 (2015)

PALAEONTOLOGY

Maternal care evolved early

Fossils of a female crustacean — the oldest known example of a female animal with eggs — suggest that parental care is almost as ancient as animals themselves.

Jean-Bernard Caron at the Royal Ontario Museum in Toronto, Canada, and Jean Vannier of Claude Bernard University Lyon in France

report the discovery of 5 well-preserved, 508-million-year-old fossils of the extinct crustacean *Waptia fieldensis*, with remnants of embryos visible. The specimens showed that *Waptia* carried broods of around 24 large eggs, each measuring up to 2.5 millimetres across, in a crevice between the body and the shell. The shell may have helped parental care to evolve by providing a safe environment to incubate eggs.

The findings suggest that parental care appeared less than 50 million years after the evolution of animals.

Curr. Biol. <http://doi.org/989> (2015)



ANIMAL BEHAVIOUR

Female elephants inherit social roles

Female elephants fill their mothers' social roles after a matriarch dies, making pachyderm networks resilient to the effects of poaching.

African elephants (*Loxodonta africana*) are organized in family groups of females, which are linked together to form 'bond' groups and loosely affiliated clans. Shifra Goldenberg at Colorado State University in Fort Collins and her team analysed the animals' female social networks in Kenyan reserves over 16 years.

There was a roughly 70% turnover in adult females from poaching and natural causes, but the overall female-led social structure persisted.

The researchers found that daughters took up their mothers' positions in networks when mothers died, and emulated their patterns of contact with other elephants. This 'network resilience' is postulated by network theory but is rarely observed in nature, the authors say.

Curr. Biol. <http://doi.org/97h> (2015)

THOMAS MANGELSEN/MINDEN PICTURES/FLPA

KENDRA HAYASHI NEGREY

QUANTUM INFORMATION

Quantum security hacked by light

Ordinary light can break the security of a standard quantum method for sharing private information.

Quantum key distribution allows two people in separate locations to use the rules of the quantum world to create a secret key, which they can use to exchange encrypted messages. But Jan-Åke Larsson at the University of Linköping in Sweden and his colleagues show how to trick one standard method for making such keys. By hijacking the light source that the two parties use to create their shared key, an attacker can control sensitive detectors at either person's location, fooling a security check into believing that no one has meddled with the protocol. The attacker can then intercept messages undetected.

The authors also show how security can be recovered by performing a different test. *Sci. Adv.* 1, e1500793 (2015)

SYNTHETIC BIOLOGY

Designer cells block psoriasis

Engineered cells with synthetic gene circuits can detect and respond to disease biomarkers. These could one day help to treat psoriasis, a chronic inflammatory skin disorder.

Clinical trials have shown that proteins called cytokines can help people who have psoriasis, but the cytokines need to be administered continuously. To overcome this, Martin Fussenegger at the Swiss Federal Institute of Technology in Zurich and his colleagues designed human cells to detect TNF and IL-22, two biomarkers that are associated with flare-ups of psoriasis. When the designer cells detect threshold levels of both biomarkers, they produce the cytokines. Implanting the cells in mouse models of skin inflammation prevented acute disease, improved skin lesions

and restored normal skin. The cells were also responsive to blood samples from people with psoriasis.

Sci. Transl. Med. 7, 318ra201 (2015)

EVOLUTION

Bacteria cannot stop adapting

One of biology's longest-running experiments suggests that adaptation can be endless, even in extremely stable environments.

To test the assumption that evolution is stimulated by environmental change, Richard Lenski at Michigan State University in East Lansing and his colleagues maintained the same populations of *Escherichia coli* in a stable environment for 27 years, freezing samples every 500 generations. They found that populations consistently outcompeted their ancestors, indicating that they were becoming increasingly fit. This continued right up to the 60,000th generation, although the rate of fitness improvement slowed over time.

The results suggest that there is no upper limit to adaptation, even in simple environments. *Proc. R. Soc. B* 282, 20152292 (2015)

INFECTIOUS DISEASE

Emerging virus evolves in camels

Nearly one-fifth of camels in Saudi Arabia harbour a respiratory virus that emerged in 2012 in humans.

Middle East respiratory syndrome coronavirus (MERS-CoV) has infected more than 1,600 people in 26 countries, killing 584. On the basis of previous evidence that camels carry the virus, a team led by Huachen Zhu and Yi Guan at the University of Hong Kong–Shenzhen Branch in China looked for MERS-CoV and related viruses in 1,309 dromedary camels in Saudi Arabia. One in four

SOCIAL SELECTION

Popular topics on social media

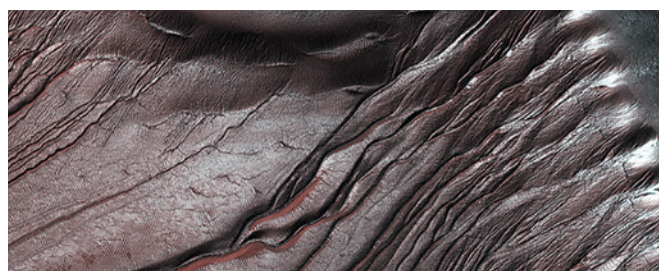
Most-tweeted papers of 2015

The hottest papers of 2015 covered topics ranging from cancer risk to reproducibility in science, according to Altmetric, a London-based company that tracks the media attention received by academic publications. In a paper ranked 9th in Altmetric's annual top 100 list, Leon Gatys at the University of Tübingen, Germany, and his team developed an algorithm that extracts and combines the content of one image with the style of another — turning a photograph into an approximation of a painting by Vincent Van Gogh, for example. The authors write that the algorithm may help to decode how humans create and perceive art. They made the model publicly available,

inspiring others such as Kai Sheng Tai, a data scientist at MetaMind in Palo Alto, California, to create their own versions of the program.

NATURE.COM
For more on popular papers:
go.nature.com/1r8hkj

<http://arxiv.org/abs/1508.06576> (2015)



camels tested positive for human coronavirus genetic material, and nearly 20% carried a MERS-CoV strain. Some animals carried the lineage that caused a South Korean outbreak this year. Further genome sequencing suggested that this lineage emerged in camels between December 2013 and June 2014, after two viruses recombined.

Preventing camel-to-human transmission is the best way to limit the threat of the virus, the authors say.

Science <http://dx.doi.org/10.1126/science.aac8608> (2015)

PLANETARY SCIENCE

No water needed for Mars gullies

Gullies on Mars can be formed by dry carbon dioxide and do not need liquid water.

Planetary scientists have been excited about gullies on Mars's surface (**pictured**)

because they look like they could have been formed recently by flowing water — possibly making the planet habitable. Cedric Pilorget at Paris-Sud University and François Forget of the Sorbonne Universities in Paris used a numerical model to simulate a layer of CO₂ ice sitting on top of the Martian soil and in pores within it. They calculated that as the Martian winter turns to spring, the ice turns to gas, destabilizing the surface and causing it to crumble and form the gullies.

The work bolsters the idea that many Martian landforms can be created by dry geological processes that do not require water.

Nature Geosci. <http://dx.doi.org/10.1038/ngeo2619> (2015)

NATURE.COM
For the latest research published by Nature visit:
www.nature.com/latestresearch

SEVEN DAYS

The news in brief

FUNDING

US science funding

A budget bill passed by the US House of Representatives on 18 December gave big spending boosts to several US science agencies for the 2016 fiscal year. The National Institutes of Health was widely hailed as the biggest winner, receiving a 6.6% increase over last year's budget, to US\$32.1 billion. NASA also fared well, with an extra \$1.3 billion, raising its total funding to \$19.3 billion. The National Science Foundation saw smaller gains: its budget of \$7.5 billion is just 1.6% more than last year's. See page 446 and go.nature.com/qxwnce for more.

EVENTS

Olive-tree disease

Nine scientists and an official in Italy are being investigated in connection with the outbreak of a bacterial disease that is ravaging the region's olive groves. On 18 December, public prosecutors announced the formal investigation, and halted a cull of 2,000 infected trees. Prosecutors cite concerns that the bacterium, *Xylella fastidiosa*, may have escaped into the environment after being imported from California for a workshop at

NUMBER CRUNCH

+1.3 °C

The average air temperature anomaly — difference from the historical average — over land in the Arctic from October 2014 to September 2015, the highest since 1900.

Source: Arctic Report Card, US National Oceanic and Atmospheric Administration



CHEN YEHUA/XINHUA/CORBIS

Postdoc dies in chemistry-lab fire

A postdoctoral researcher died following an explosion on 18 December in the chemistry department of Tsinghua University in Beijing. According to a notice on the university's official Weibo social-media account, which confirmed the researcher's death, the explosion occurred at 10.10 a.m. local time. The university stated

on Weibo that the fire had been extinguished, and that other personnel had been evacuated. Images shared on social media showed black smoke billowing out of a window of the red-brick Ho Tim building. The cause of the event is under investigation. See go.nature.com/x4myfb for more.

the Mediterranean Agronomic Institute of Bari in 2010. The scientists deny using the *X. fastidiosa* strain in question in the workshop. See go.nature.com/8ejnby for more.

Endocrine ruling

The European Commission acted unlawfully in failing to design scientific procedures to identify chemicals that may affect hormone levels in humans as part of 2012 legislation, the European Court of Justice ruled on 16 December. Sweden brought the case in 2014 after the commission failed to establish criteria for detecting suspected 'endocrine disrupter' chemicals such as bisphenol A — found widely in food, plastics and cleaning products — as

required by the legislation. Some scientists say that the chemicals harm health. The commission, which has two months to appeal the decision, will complete an ongoing impact assessment of endocrine disrupters in 2016 and will establish detection criteria thereafter.

Space success

The California-based space-flight company SpaceX has for the first time soft-landed a rocket booster after using it to propel a payload into orbit. The second stage of the Falcon 9 vehicle, which lifted off on 21 December from Cape Canaveral, Florida, deployed 11 satellites into orbit; the first stage returned to Earth at a landing site 10 kilometres

down the coast, using its boosters to slow its descent. The company had made two previous attempts at rocket retrieval earlier this year, both of which resulted in crashes.

BUSINESS

Viral cancer drug

The European Commission on 17 December approved a trail-blazing cancer-fighting virus called talimogene laherparepvec for the treatment of advanced melanoma. The virus — a modified live herpesvirus made by biotechnology giant Amgen of Thousand Oaks, California — destroys cancer cells directly while also triggering an immune response. The US Food and Drug Administration

approved the drug — the first of its kind to hit the market — on 27 October. See go.nature.com/wlyee for more.

PEOPLE

Harassment report

The University of California, Berkeley, has released its report on sexual-harassment complaints against astronomer Geoff Marcy, who in October stepped down from his faculty position after the accusations came to light. The university made the report and related documents available on 17 December, in response to public-records requests. The documents detail the university's investigation, which ultimately involved four harassment complaints from four individuals. Marcy has not publicly addressed the complaints specifically, and neither he nor his lawyer has responded to *Nature's* request for comment. See go.nature.com/iddc7i for more.

Fraud charges

A controversial pharmaceutical-company executive resigned from one of his posts on 18 December, after being indicted by the US justice department. Martin Shkreli (pictured, centre) stepped down from his role at the company he founded, Turing Pharmaceuticals of New York City, a day after he



was charged with securities fraud in connection with two hedge funds and a drug company he used to run. Shkreli gained notoriety earlier this year when Turing increased the price of an anti-parasite drug from US\$13.50 to \$750 per pill.

POLICY

EU catch limits

The European Union was criticized last week after setting limits for fish catches that were higher than those recommended by scientific advice. On 16 December, ministers from member states agreed on how much fish could be caught in the Atlantic Ocean, North Sea and Black Sea in 2016. But non-governmental organizations and some researchers say that the limits set on some fish stocks exceed levels recommended by independent scientists, and that this

threatens the EU's aim of fishing sustainably by 2020. See page 435 for more.

US climate vetoes

US President Barack Obama has blocked two bills, both approved by Congress, that would have voided regulations to limit greenhouse-gas emissions from power plants. "Climate change poses a profound threat to our future and future generations," Obama wrote in a 18 December letter to Congress announcing his veto of one of the bills. The president vetoed the second bill on 19 December, outlining the decision in a separate letter to Congress.

Lion protection

Lions in Africa and India will receive protections under the US Endangered Species Act, the US Fish and Wildlife Service said on 21 December. Populations in India and West

and Central Africa will be listed as endangered, and lions in southern and eastern Africa will be classed as threatened. Under the designation, US hunters will not be allowed to import lion trophies to the United States under most circumstances. The move comes five years after several conservation groups called on the US government to deem African lions endangered.

US chemical reform

On 17 December, the US Senate passed an update to the Toxic Substances Control Act, a 1976 law that gives the Environmental Protection Agency (EPA) authority to regulate chemicals used in consumer goods and industry. Unlike the existing law, the updated bill would not allow a new chemical to come to market unless the EPA found it likely to be safe. The House of Representatives passed a similar law in June, and the two houses will now attempt to resolve their differences before voting on a single bill and sending it to the president. See go.nature.com/sfug44 for more.

Data exemptions

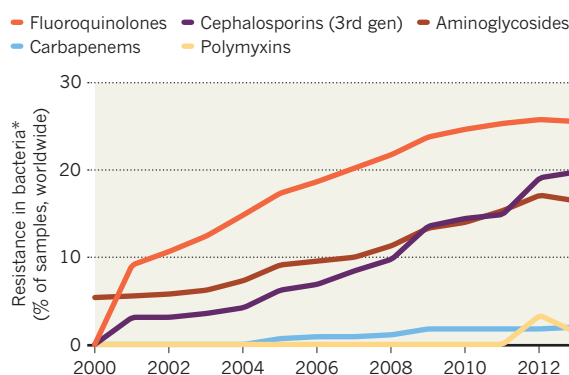
European Union politicians and officials agreed on 15 December to exempt scientific research from certain regulations in planned data-protection legislation. Among other laws, research will be exempted from a rule that all personal data remain anonymous indefinitely, which would make it hard for medical researchers to track long-term disease progression. Another rule would have required researchers to obtain fresh consent from donors every time their data or tissues were used in a different study. The compromise allows medical researchers to unmask data in special circumstances and to reuse data and samples for multiple studies in different diseases, as long as a general consent form is signed.

TREND WATCH

Researchers have found that bacteria worldwide share a gene that confers resistance to colistin, a 'last resort' antibiotic. Discovery of the gene was reported in China last month, and has been followed by findings of similar resistance in countries including Denmark, France and Thailand. Bacteria have been slow to develop resistance to colistin — a polymyxin antibiotic developed in the 1950s — compared with other antibiotics because it is little-used in humans. See go.nature.com/hbh2qe for more.

THE SPREAD OF ANTIBIOTIC RESISTANCE

An increasing proportion of bacteria display resistance to common antibiotics.



*Enterobacteriaceae, including *Escherichia coli*, *Klebsella pneumoniae*, *Enterobacter* and *Salmonella*

NATURE.COM

For daily news updates see:

www.nature.com/news

NEWS IN FOCUS

POLITICS Canada's first science minister brings air of change **p.445**

STRING THEORY Philosophers join debate over scientific method **p.446**



2015 IN REVIEW Gene editing, climate change, Pluto and more **p.448**

PROFILES Ten people who mattered in science this year **p.459**

JIN LIWANG/XINHUA/EYEVINE



The Monkey King spacecraft, which took to the skies on 17 December, is designed to detect the high-energy particles produced by annihilating dark matter.

COSMOLOGY

Dark-matter probe launches era of Chinese space science

Monkey King is first in a line of Chinese space missions focused on scientific discovery.

BY ELIZABETH GIBNEY, CELESTE BIEVER & DAVIDE CASTELVECCHI

Against a purple morning sky, in a cloud of brown smoke, the Monkey King took off. China's first space-based dark-matter detector — nicknamed Wukong (or Monkey King) after a warrior in a sixteenth-century Chinese novel — rocketed into the air on 17 December, marking the start of a new direction in the country's space strategy.

From Earth's orbit, the craft aims to detect

high-energy particles and γ -rays. Physicists think that dark matter — a substance thought to make up 85% of the Universe's matter but so far observed only through its gravitational effects — could reveal itself by producing such cosmic rays as its constituent particles annihilate.

Wukong, officially called the Dark Matter Particle Explorer (DAMPE), is also notable for being the first in a series of five space-science missions to emerge from the Chinese Academy of Sciences' Strategic Priority Program on Space Science, which kicked off in 2011.

China is already one of the world's major space powers, but so far has focused on human and robotic exploration, with little investment in space science. (A notable exception is the Double Star probe launched in collaboration with the European Space Agency in 2003 to study magnetic storms on Earth.)

The DAMPE lift-off from the Jiuquan Satellite Launch Center in northern China will be followed next year by a further two missions: the world's first quantum-communications satellite and an X-ray telescope observing in ►

► a unique energy band. Together, these missions mark a new start for space science in China, says Wu Ji, director-general of the National Space Science Center (NSSC), which runs the programme.

Other countries have had Moon missions, adds Pan Jian-Wei, chief scientist for the quantum-science satellite, but with the space-science satellites, “we can do something new and something really great and not only for China — for the whole world”.

The public gave DAMPE its nickname, Wukong, earlier this week, as part of an outreach drive in China’s space programme; a similar open effort also produced the name Yutu, or Jade Rabbit, for the nation’s lunar rover, which landed in 2013.

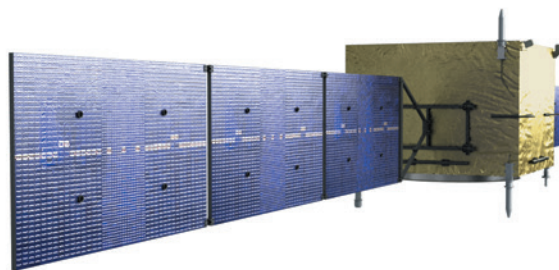
Wukong will use its relatively large detection area to observe high volumes of cosmic rays, as well as where they come from. It will survey the sky at energies much higher than do existing detectors such as the Alpha Magnetic Spectrometer (AMS), which is currently attached to the International Space Station. “We don’t know if this is a better way to search for dark matter, because dark matter has not yet been found,” says Michael Capell, an AMS physicist at CERN, Europe’s particle-physics laboratory near Geneva, Switzerland.

PULSAR PUZZLE

The detector could help to clear up some mysteries. In 2013, the AMS team announced that it had seen hints of dark matter, but so far it has detected too few high-energy particles to say for sure. DAMPE lacks the equipment to clarify the situation directly, but it could reveal whether the signal is from an astrophysical source other than dark matter, such as pulsars, says Capell.

Although it will collect fewer incoming photons than existing γ -ray telescopes such as NASA’s Fermi-LAT, DAMPE is better at pinpointing the energies of these particles, says Miguel Sánchez-Conde, a physicist at the Oskar Klein Centre for Cosmoparticle Physics in Stockholm. This capability should allow DAMPE to see sharp spikes in radiation that are predicted by some dark-matter models.

The two experiments that will follow hot on Wukong’s heels are no less ambitious. The quantum-science satellite, to launch in June,



China’s DAMPE probe.

will be the world’s first space experiment to probe the phenomenon known as quantum entanglement. The mission will test whether a pair of entangled photons beamed from the satellite to two ground stations can remain entangled over a record-breaking distance of more than 1,000 kilometres.

The experiment will also test whether a quantum connection can be set up between a ground station and the satellite and used to ‘teleport’ information instantly and securely. Previously, such experiments have transmitted photons on Earth through optic fibres or air, and over much shorter distances. The eventual goal is to create a global quantum-communications network, says Anton

“We can do something new and something really great and not only for China.”

Zeidler, a physicist at the University of Vienna who is collaborating with Pan on the quantum satellite. By pushing the limits of quantum entanglement, Pan says, the satellite may also help to solve fundamental mysteries about the Universe, such as how to unite quantum mechanics with Einstein’s general theory of relativity.

BLACK HOLES

In the second half of the year, China will launch the Hard X-ray Modulation Telescope (HXMT), looking for bright and brief sources of radiation, such as growing black holes. The HXMT will do a broad sweep of the sky, with a sensitivity at the top of its large energy range that exceeds those of existing wide-field telescopes, says Luigi Piro, an astronomer at Italy’s National Institute for Astrophysics in Rome.

All three are cutting-edge missions, with the potential to make real discoveries, says

Wu — but he is still not satisfied. Space science in China is funded in 5-year cycles, receiving around 3 billion yuan (US\$460 million) in the current round. As a result there is no permanent funding, unlike in the United States and Europe, which makes it difficult to make long-term plans. “We don’t feel it is secure,” says Wu. “It is better than nothing. But we are still catching up.” He believes that until China

makes discoveries in space science, “we are not a real space power”.

The current funding round runs out next year. Although Wu thinks that the Chinese Academy of Sciences will continue to support the programme for another five years, that will be confirmed only next year. The funding will have to cover the remaining two missions — a satellite, Shijian-10, to conduct microgravity and life-sciences experiments, and a space-weather satellite known as Kuafu.

INTERNATIONAL CONTRIBUTIONS

Piro notes that most of the present and future Chinese scientific satellites include research contributions from scientists worldwide. Such collaborations “sharpen scientific goals, optimize resources and avoid overlap”, he says. Zeidler attributes China’s pioneering work in space-based quantum communications to fast decision-making processes “oriented towards getting things done”.

The US Congress passed a law in 2011 that prevents NASA from collaborating with China except in rare circumstances. By contrast, the European Space Agency wants to work with China and is already collaborating with the Chinese academy on a small space-weather observatory, the Solar wind Magnetosphere Ionosphere Link Explorer (SMILE).

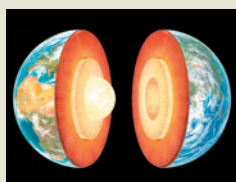
China’s limited experience in space science, alongside its politics, has hampered collaboration so far, says Joan Johnson-Freese, who specializes in China’s space programme at the US Naval War College in Newport, Rhode Island. But the country is anxious to develop and establish its expertise, she adds.

Chinese scientists would like to collaborate with the United States, says Wu, but the severed ties hurt the United States more than China. “It gave a good chance for the Europeans. The US should realize that.” ■



**MORE
ONLINE**

TOP NEWS



Magnetic mystery of Earth’s early core explained go.nature.com/sahnnx

MORE ONLINE

- Atom-thin ‘borophene’ joins 2D materials club go.nature.com/nydi7a
- Biggest mystery in mathematics in limbo after cryptic meeting go.nature.com/xbfktt
- LHC sees hint of boson heavier than Higgs go.nature.com/iqftek

NATURE PODCAST



The year’s science highlights, busting scientific myths, and an evolution-themed board game nature.com/nature/podcast



Kirsty Duncan was first elected to Canada's Parliament in 2008.

POLITICS

Canada's top scientist faces tough challenge

Researchers have big hopes for Kirsty Duncan, the country's newly appointed scientist-turned-science minister.

BY NICOLA JONES

Kirsty Duncan, the medical geographer who last month became Canada's first Minister of Science, has a big mandate: to ensure that scientific considerations again figure into public-policy decisions.

Duncan, appointed by newly elected Prime Minister Justin Trudeau, inherits a research community bruised by years of cuts to science programmes and research jobs under former prime minister Stephen Harper. Harper's government also famously muzzled government researchers. But change is in the air.

On 5 November, Trudeau's government reinstated the mandatory long-form census, to cheers from social scientists, and on 6 November, it decreed that federal scientists could again speak freely to the media and to the public.

Yet these splashy announcements came not from Duncan, but from Navdeep Bains, the Minister of Innovation, Science and Economic Development. Duncan has harder tasks ahead, says Paul Dufour, a science-policy analyst at the University of Ottawa. She has been asked to shoulder the burden of shoring up Canada's science enterprise; this includes steps such as reforming the country's weakened environmental-assessment

process and making basic research a higher funding priority.

But it is not clear whether she will have the power to make such changes. Canada's science ministers have historically operated with minimal budgets, and sometimes as junior ministers. Duncan's clout will not be put to the test until Trudeau releases his first federal budget in February. "She's a great person for the job, but is it window dressing?" says Kennedy Stewart, who tracks science issues for the New Democratic Party, the left-wing opposition to Duncan and Trudeau's middle-left Liberal party. "The budget will tell."

In Canada, where ministers are chosen from among elected members of parliament, it is rare to see higher degrees in fields other than law or medicine. Trudeau's cabinet is a notable exception: Duncan, who earned a PhD in geography in 1992 at the University of Edinburgh, UK, is one of a small group of ministers with doctoral degrees in economics, sociology or engineering.

Duncan is perhaps best known for leading an expedition to Norway in 1998, prompted

by her interest in pandemics. Then at Canada's University of Windsor, she suspected that traces of the deadly 1918 Spanish flu virus might be preserved in the bodies of victims who were buried in permafrost.

Although the expedition did not yield any flu samples, team member Robert Webster, a pandemic virologist at St Jude Children's Research Hospital in Memphis, Tennessee, remains impressed by Duncan's organizational acumen. "She was smart enough to contact the leaders in the field," he says. "She got the heavies. She raised the funds."

Economist Paul Kovacs, who worked with Duncan on a chapter of the Intergovernmental Panel on Climate Change's 2001 report, makes a similar assessment. Kovacs, executive director of the Institute for Catastrophic Loss Reduction at the University of Western Ontario in London, Canada, describes her as dedicated, determined and skilled at probing the scientific literature to work out "what was really new and what you could do about it".

TRIAL TRIBULATIONS

But Duncan's political career, which began in 2008, has not been without controversy. Between 2012 and 2014 she introduced seven pieces of legislation, all related to neurological health. Two bills called for clinical trials of controversial treatments for multiple sclerosis; these were based on the work of Paolo Zamboni, an Italian physician who suggested that a circulatory condition called chronic cerebrospinal venous insufficiency was linked to the neurological disorder. Duncan's bills came after several studies failed to find evidence for Zamboni's claims, and concluded that the therapy was too expensive and risky for further trials.

But Duncan defends the legislation, saying that she wanted to encourage research on the brain. "In science we ask the questions. I asked a question: would the government look at the science?" she says.

In the long term, Duncan will work to improve Canada's science capacity — in part by establishing high-profile professorships in sustainable technologies. According to the United Nations, the country is one of only a few advanced economies whose total spending on research and development has declined relative to its gross domestic product.

Observers are keen to see what Duncan can achieve. "She's certainly got her hands full with limited resources," Dufour says.

For now, Duncan is focused on establishing the post of chief science officer, to replace the national science adviser role that Harper eliminated in 2008. Physicist Ted Hsu, the Liberal party's former science spokesperson, says that this will take some thought. "She needs to set up something that's so good, it will survive a change of government in future."

Duncan is happy to go slowly to work out the best system. "We want to get this right," she says. ■

FUNDING

US biomedicine nets budget win

Late spending bill gives the NIH healthy increases.

BY SARA REARDON, CHRIS CESARE & HEIDI LEDFORD

Biomedical research advocates are revelling in holiday cheer as a budget bill passed by Congress and signed into law by President Obama on 18 December gives the US National Institutes of Health (NIH) its biggest funding increase since 2003. Several other science-related agencies also benefit substantially from the budget.

“Best Christmas present ever,” says Jennifer Zeitzer, director of legislative relations at the Federation of American Societies for Experimental Biology in Bethesda, Maryland. The budget allocates just over US\$32.1 billion to the agency: a 6.6% rise over 2015. Accounting for inflation, the agency’s funding had fallen 20% compared with 2003; the new budget, Zeitzer says, almost returns the NIH to its real 2003 level.

Several other research agencies have found similarly generous gifts in the budget, which was approved 11 weeks after the 1 October beginning of the 2016 fiscal year. NASA gets a bump of almost \$1.3 billion over its 2015 funding, to \$19.3 billion. That sum includes \$175 million for a mission that will orbit and land on Jupiter’s icy moon Europa and search for signs of life.

The budget allocates \$7.5 billion to the National Science Foundation (NSF), a small 1.6% increase over 2015 levels. The document does little to specify how the NSF spends its money — a contentious issue that arose in June when the Republican-controlled House of Representatives proposed requiring the foundation to spend 70% of its research funds on biology, computer science, engineering, mathematics and physical sciences. The provision would have effectively cut the funds available to social science and geoscience by about 15%. In the end, the spending bill specifies only that social-sciences spending remain flat.

Although the healthy funding increases come as good news to many researchers, says Michael Lubell, the director of public affairs at the American Physical Society in Washington DC, there is bad news on the horizon. He points out that a deal struck in October by legislators and Obama provides almost no room for further boosts in 2017. “One should not say all of this is ushering in a new era,” Lubell says. “It is not.” ■



R. WINDHORST, ARIZONA STATE UNIV./H. YAN, SPITZER SCIENCE CENTER, CALTECH/ESA/NASA

The idea that our Universe is part of a multiverse poses a challenge to philosophers of science.

COSMOLOGY

Feuding physicists turn to philosophy

String theory is at the heart of a debate over the integrity of the scientific method itself.

BY DAVIDE CASTELVECCHI

Is string theory science? Physicists and cosmologists have been debating the question for the past decade. Now the community is looking to philosophy for help.

Earlier this month, some of the feuding physicists met with philosophers of science at an unusual workshop aimed at addressing the accusation that branches of theoretical physics have become detached from the realities of experimental science. At stake is the integrity of the scientific method, as well as the reputation of science among the general public, say the workshop’s organizers.

Held at the Ludwig Maximilian University of Munich in Germany on 7–9 December, the workshop came about as a result of an article in *Nature* a year ago, in which cosmologist George Ellis, of the University of Cape Town in South Africa, and astronomer Joseph Silk, of Johns Hopkins University in Baltimore, Maryland, lamented a “worrying turn” in theoretical physics (G. Ellis and J. Silk *Nature* 516, 321–323; 2014).

“Faced with difficulties in applying fundamental theories to the observed Universe,” they wrote, some scientists argue that “if a theory is sufficiently elegant and explanatory,

it need not be tested experimentally”.

First among the topics discussed was testability. For a scientific theory to be considered valid, scientists often require that there be an experiment that could, in principle, rule the theory out — or ‘falsify’ it, as the philosopher of science Karl Popper put it in the 1930s. In their article, Ellis and Silk pointed out that in certain areas, some theoretical physicists had strayed from this guiding principle — even arguing for it to be relaxed.

The duo cited string theory as the principal example. The theory replaces elementary particles with infinitesimally thin strings to reconcile the apparently incompatible theories that describe gravity and the quantum world. The strings are too tiny to detect using today’s technology — but some argue that string theory is worth pursuing whether or not experiments will ever be able to measure its effects, simply because it seems to be the ‘right’ solution to many quandaries.

Silk and Ellis also called out another theory that seems to have abandoned ‘Popperism’: the concept of a multiverse, in which the Big Bang spawned many universes — most of which would be radically different from our own.

But in the opening talk at the workshop, David Gross, a theoretical physicist at the

University of California, Santa Barbara, drew a distinction between the two theories. He classified string theory as testable “in principle” and thus perfectly scientific, because the strings are potentially detectable.

Much more troubling, he says, are concepts such as the multiverse because the other universes that it postulates probably cannot be observed from our own, even in principle. “Just to argue that [string theory] is not science because it’s not testable at the moment is absurd,” says Gross, who shared a Nobel prize in 2004 for his work on the strong nuclear force, which is well tested in experiments, and has also made important contributions to string theory.

Workshop attendee Carlo Rovelli, a theoretical physicist at Aix-Marseille University in France, agrees that just because string theory is not testable now does not mean that it is not worth theorists’ time. But the main target of Ellis and Silk’s piece were observations made by philosopher Richard Dawid of Ludwig Maximilian University in his book *String Theory and the Scientific Method* (Cambridge Univ. Press, 2013). Dawid wrote that string theorists had started to follow the principles of Bayesian statistics, which estimates the likelihood of a certain prediction being true on the basis of prior knowledge, and later revises that estimate as more knowledge is acquired. But, Dawid notes, physicists have begun to use purely

theoretical factors, such as the internal consistency of a theory or the absence of credible alternatives, to update estimates, instead of basing those revisions on actual data.

DYNAMIC DISCUSSION

At the workshop, Gross, who has suggested that a lack of alternatives to string theory makes it more likely to be correct, sparred with

“Suggestions that we need ‘new methods’ have been made, but attempts to replace empirical testability have always failed.”

Rovelli, who has worked for years on an alternative called loop quantum gravity. Rovelli flatly opposes the assumption that there are no viable alternatives. Ellis, meanwhile, rejects the idea that theo-

retical factors can improve odds. “My response to Bayesianism is: new evidence must be experimental evidence,” he says.

Others flagged up separate issues surrounding the use of Bayesian statistics to bolster string theory. Sabine Hossenfelder, a physicist at the Nordic Institute for Theoretical Physics in Stockholm, said that the theory’s popularity may have contributed to the impression that it is the only game in town. But string theory probably gained momentum for sociological

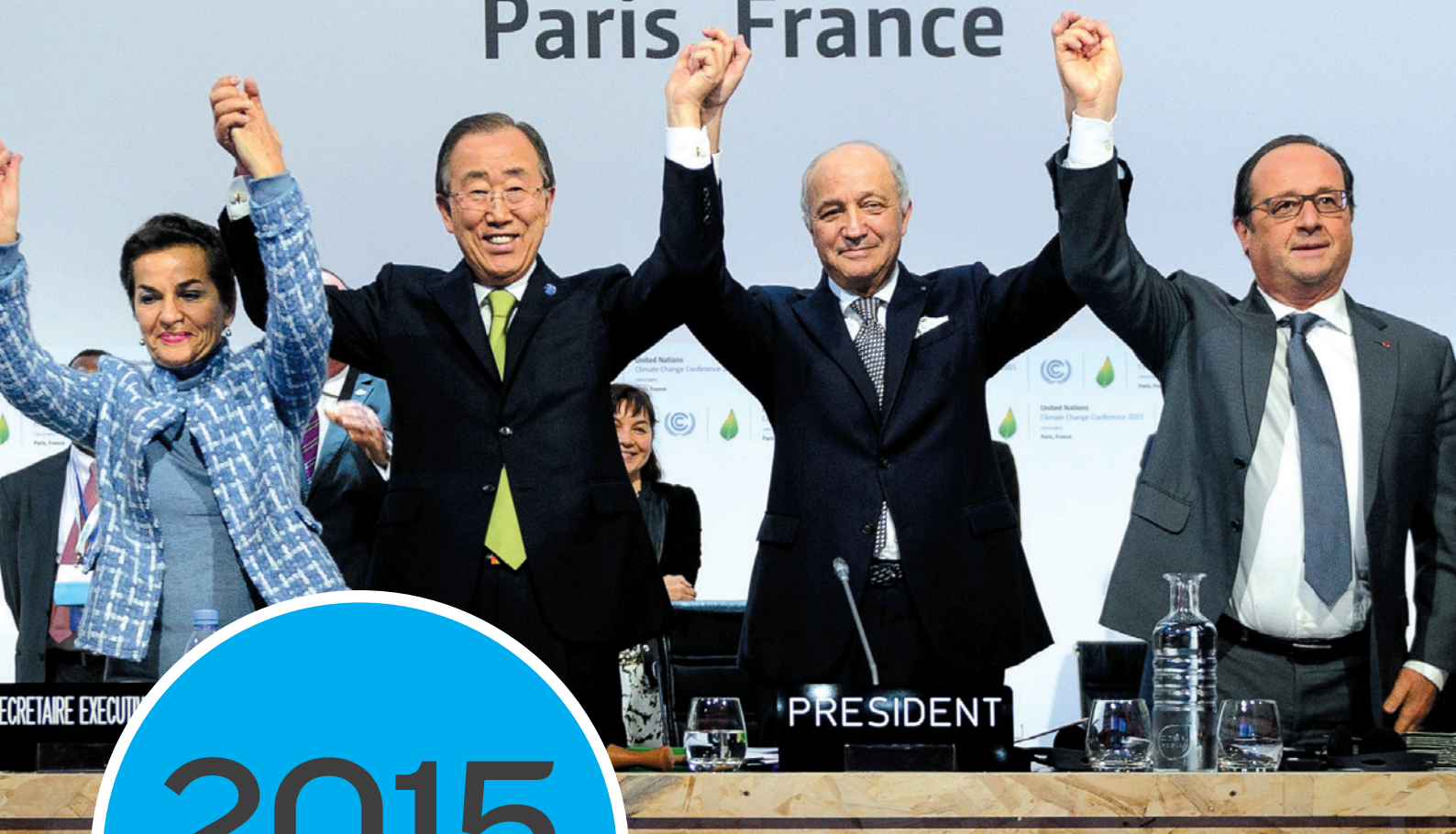
reasons, she said: young researchers may have turned to it because the job prospects are better than in a lesser-known field, for example.

Historian of science Helge Kragh of Aarhus University in Denmark drew on historical perspective. “Suggestions that we need ‘new methods of science’ have been made before, but attempts to replace empirical testability with some other criteria have always failed,” he said. But at least the problem is confined to just a few areas of physics, he added. “String theory and multiverse cosmology are but a very small part of what most physicists do.”

That is cold comfort to Rovelli, who stressed the need for a clear distinction between scientific theories that are well established by experiments and those that are speculative. “It’s very bad when people stop you in the street and say, ‘Did you know that the world is made of strings and that there are parallel worlds?’”

At the end of the workshop, the feuding physicists did not seem any closer to agreement. Dawid — who co-organized the event with Silk, Ellis and others — says that he does not expect people to change their positions in a fundamental way. But he hopes that exposure to other lines of reasoning might “result in slight rapprochement”. Ellis suggests that a more immersive format, such as a two-week summer school, might be more successful at producing a consensus. ■

Paris France



2015 IN REVIEW

From climate change to gene-editing ethics, researchers tackled many thorny issues this year. They also made important discoveries — including ice mountains on Pluto, evidence of quantum weirdness and details about the molecular machines inside cells.

ROAD TO PARIS The world got serious this year about climate change. With the United Nations climate summit in Paris looming in December, both industrialized and developing nations pledged for the first time to control or reduce their greenhouse-gas emissions.

As the number of pledges grew during the year — to 184 by the time of the conference — so did optimism that the Paris talks would be a historic turning point in efforts to curb global warming (see page 460). The meeting, which took place under heightened security because of the Paris terrorist attacks in November, yielded a landmark agreement on 12 December that was approved by 195 countries. It commits most countries to reduce emissions and keep warming to 'well below' 2°C. Nations will assess their progress in 2018 and must revisit their climate pledges every five years starting in 2020.

Climate negotiators were treated to some surprising good news in early December, when researchers at the Global Carbon Project reported that global carbon emissions could drop by 0.6% in 2015.

China and the United States, the world's biggest carbon emitters, helped to build momentum in the run-up to Paris. China announced that it would launch an emissions cap-and-trade system. And after years of indecision, US President Barack Obama made the symbolic move of saying no to the Keystone XL pipeline that would have transported oil from Canada to US refineries.

Even Pope Francis weighed in. He released an encyclical on the environment in June and gave speeches during his visit to North America in September that warned of the dangers of climate change and the urgent need to curb

it. Two surveys of people in the United States that were conducted after the Pope's visit suggested that he helped to boost the acceptance of climate change as an important problem.

But nations' climate pledges will probably not keep warming to within 2°C above pre-industrial levels, and past that point, many scientists think that the world will see warming-related ecological and economic disruptions. The average global surface temperature is now already 1°C above pre-industrial levels, and 2015 will probably be the warmest year on record.

PLUTO ET AL. In Solar System exploration, dwarf planets ruled. The tiny worlds of Pluto and Ceres — the latter in the heart of the asteroid belt between Mars and Jupiter — received their first-ever spacecraft visits in

ARNAUD BOISSOU/COP21/ANADOLU/GETTY

Leaders at the UN climate meeting in Paris celebrated the adoption of a historic global warming agreement on 12 December.

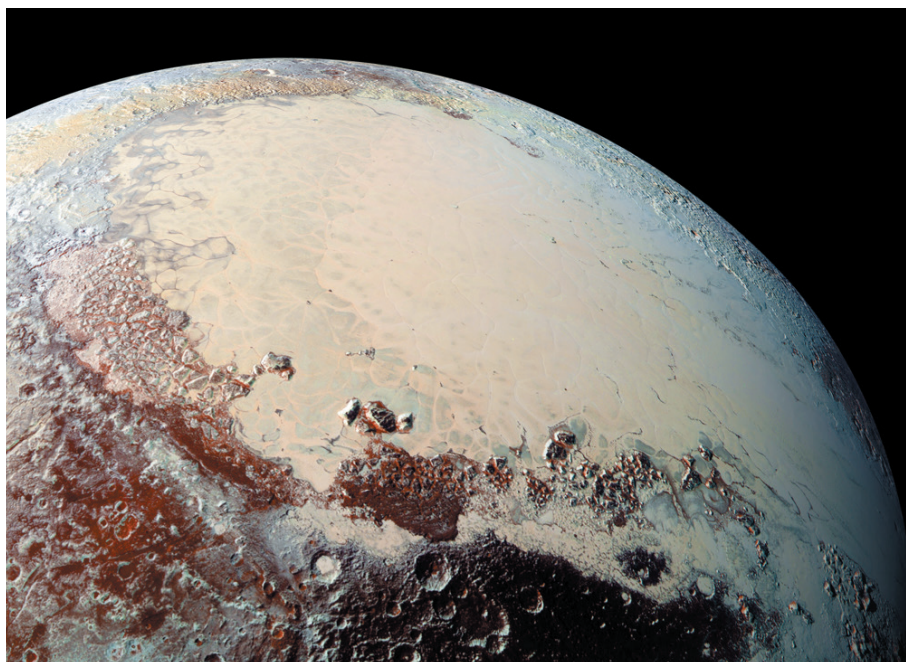
2015, producing many breathtaking images.

Pluto grabbed the spotlight when the New Horizons spacecraft flew past it on 14 July. The world revealed itself as a geological wonderland of ice mountains, nitrogen glaciers and smooth, frigid plains. The sheer complexity of Pluto's surface astounded planetary scientists, including principal investigator Alan Stern (see page 462), and raised major questions about what could be fuelling the geological activity that created it.

Ceres made a much more gradual appearance beginning in March, when its gravitational pull tugged NASA's Dawn spacecraft into orbit. The dark, water-rich body turns out to hold a number of its own mysteries, including a pyramid-shaped mountain, bright spots of reflective salt and an enigmatic haze that fills some of its craters in the morning sunlight.

The European Space Agency's Rosetta craft continued its spectacular orbit around Comet 67P/Churyumov–Gerasimenko. Its Philae lander, presumed lost after a bumpy landing in November 2014, phoned home in June before falling silent, perhaps permanently, the following month. Researchers analysing Rosetta data reported this year that oxygen is streaming out of the comet, and that its rubber-duck shape was probably a result of a low-speed collision between two smaller comets.

NASA's MAVEN (Mars Atmosphere and Volatile Evolution) mission delivered its first detailed measurements of how the solar wind strips away Mars's atmosphere over time, leading to the mostly airless world that Mars is today. And 11 years after arriving at the Saturn system, NASA's Cassini spacecraft confirmed that the buried ocean beneath the surface of the moon Enceladus stretches around the entire globe — making it a tempting place to hunt for extraterrestrial life.



NASA's New Horizons spacecraft sent back spectacular images of Pluto's rich terrain.

GENE EDITS TO ORDER Rarely has a method roared onto the scene as quickly as the accurate, easy-to-use yet controversial CRISPR–Cas9 genome-editing system. In April, scientists in China reported use of the technique to edit non-viable human embryos (see page 461), which spurred researchers and bioethicists to debate in editorials and meetings whether the technology should ever be used in human embryos, even for basic research. The debate culminated in the International Summit on Human Gene Editing in early December in Washington DC, which brought together nearly 500 ethicists, scientists and legal experts from more than 20 countries. The organizers wrapped up the event with a statement: the tools are not yet ready to be used to edit the genomes of human embryos intended for pregnancy. But they did

not call for an outright ban of this work for basic research.

Over the past three years, CRISPR has become the tool of choice for scientists seeking to enhance animals and crops, and to cure human disease (see 'CRISPR craze'). In October, researchers set a record by editing the genomes of pig embryos in 62 places at once — a move that could help to revitalize the field of xenotransplantation. The genetic tinkering could lower the risk of exposure to potentially dangerous pig viruses when people receive human-like organs grown in swine. Dogs, goats and sheep have also had their DNA modified with the low-cost technology.

CRISPR could target human diseases as well. With that aim in mind, in August, Google and other investors pumped US\$120 million into the genome-editing start-up Editas Medicine in Cambridge, Massachusetts. The firm plans to use CRISPR in clinical trials in 2017 to correct a genetic mutation in some people who are visually impaired.

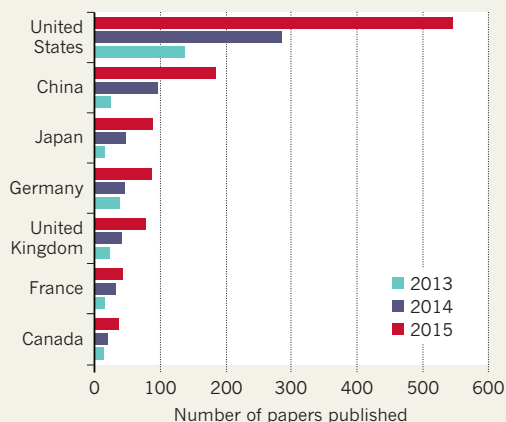
Other, more mature genome-editing technologies are already entering the clinic. In November, researchers in the United Kingdom announced that they had used a different system — enzymes called TALENs — to edit human immune cells and transplant them into a one-year-old with leukaemia, possibly saving her life.

And in December, scientists from Sangamo Biosciences in Richmond, California, announced that in 2016 they will begin a human trial to test DNA-snipping zinc-finger nucleases that correct a gene defect for haemophilia. ▶

SOURCE: SCOPUS; IMAGE: BGI

CRISPR CRAZE

Research using the CRISPR gene-editing system is ramping up, as seen by the rise in the number of CRISPR-related publications.



A technician from Chinese genomics institute BGI holds a 'micropig', whose genome was edited using TALEN enzymes.



► **VACCINE VICTORIES** Edward Jenner, who tested the first vaccine more than 200 years ago, would have been proud of the progress in 2015. After being fast-tracked into human trials this April, the rVSV-ZEBOV Ebola vaccine was found to offer near-total protection to people who received it soon after exposure to the disease, according to preliminary analysis of an ongoing clinical trial in Guinea. The vaccine consists of a weakened livestock virus that has been engineered to produce an Ebola protein, and it was the result of an accelerated development programme that experts say could be emulated to combat other emerging diseases.

But rVSV-ZEBOV arrived too late to have much impact on the Ebola epidemic, which has killed more than 11,000 people across West Africa. The disease is on the wane, but it made a surprising comeback in Liberia recently; after twice saying that it had rid itself of the virus, the country announced three new cases in November, including one death.

Nearly 30 years in the making, the world's first malaria vaccine won a lukewarm endorsement from a global vaccine advisory group in October. Researchers reported in April that the vaccine achieved a modest 30% protection rate in a clinical trial involving more than 15,000 children in Africa. The panel recommended pilot tests of the vaccine, called RTS,S, in up to 1 million children before it is widely distributed.

Polio vaccines brought the debilitating disease nearer than ever to global eradication: this year, just 66 wild-poliovirus cases were recorded as of 9 December. In July, Nigeria — one of three countries, along with Pakistan and Afghanistan, that have never interrupted the spread of the virus — celebrated a full year without a new wild-poliovirus infection for the first time, prompting the World Health Organization to remove the

5,154

A RECORD NUMBER OF
AUTHORS ON A PAPER
WAS SET THIS YEAR.

country from its list of polio-endemic nations in September. This paves the way for Africa to be declared polio-free as early as 2017.

Finally, Mexico approved the first ever vaccine against dengue virus. The vaccine's maker, Paris-based Sanofi, now hopes to secure approval in other countries in Latin America and Asia.

QUANTUM SPOOKINESS Physicists celebrated the 100th anniversary of Albert Einstein's general theory of relativity in November with special conferences, books and collections of his papers. Einstein also made headlines in August when physicists presented the most convincing proof yet that two objects, such as subatomic particles, could be linked, or 'entangled'. This would allow one particle to influence the behaviour of another, even if the two are widely separated. Researchers showed that they could produce a robust entanglement between two electrons placed 1.3 kilometres from each other.

Einstein famously despised this phenomenon, which he called 'spooky action at a

distance', because it seemingly broke the universal rule that nothing can travel faster than the speed of light. Despite Einstein's misgivings, the approach could one day be used to build a highly secure quantum Internet that is immune to hackers.

ARTIFICIAL EARTHQUAKES Oil and gas exploration and other human activities are thought to have triggered earthquakes worldwide, from Switzerland to India and China, but nowhere have scientists scrambled to understand and respond to the quakes as much as in Oklahoma. The state began recording an increase in seismicity in 2009, and this year experienced the most yet — it now has more quakes of magnitude 3 and above each year than California.

In April, officials finally acknowledged the probable role of the energy industry. The Oklahoma Geological Survey announced that oil and gas wells that pump wastewater deep into the ground are probably to blame: the injection of tens of millions of litres of liquid shifts fault stresses and increases the likelihood of quakes.

In response, the Oklahoma Corporation Commission, which regulates oil and gas exploration, cut back on the number of wastewater disposal wells allowed in the areas with the most seismic activity — a remarkable move given how powerful the energy industry is in state politics.

RESEARCH RELIABILITY RATED Debate about how to boost the reproducibility of research results shifted from handwringing to analysis and action in 2015.

Researchers in an array of fields struggle to independently reproduce published results for many reasons, ranging from poorly described methods to flawed data analysis.

In December, the US-based Reproducibility Project: Cancer Biology announced that it had scaled back its attempts to reproduce high-profile papers in cancer biology, from 50 papers to 37, because of the excessive cost and time required.

Efforts to quantify the problem bore fruit this year. In April, another Reproducibility Project team showed that some two-thirds of attempts to replicate published psychology studies ended in failure (see page 466). And a controversial analysis estimated that US\$28 billion a year is spent on biomedical studies that are not reproducible, often because of poor documentation and flawed materials.

Funders have responded. Key biomedical institutes in the United Kingdom, including the Wellcome Trust, released a report this year sketching out strategies to improve reproducibility, such as standardizing experimental practices. The US National Institutes of Health (NIH) released reproducibility guidelines in October. These asked grant reviewers to look for



A three-week-old baby in Guinea was one of the last patients in the Ebola outbreak.

SAMUEL ARANDA/NYT/REDUX/EYEVINE

PETE SOUZA/WHITE HOUSE
 flaws in experimental design that might introduce bias and requested that grant applicants describe how they will authenticate reagents. Some scientific societies pushed back this year on another set of NIH guidelines from 2014 that required authors to describe their experiments more fully. The societies said that the rules would make the preparation and reviewing of papers too burdensome. Publishers are also getting involved: around a dozen journals this year began asking their authors to use unique identifiers for their reagents as part of a push by the Resource Identification Initiative.

SPOTLIGHT ON SEXISM The discussion about sexism grew more public this year, driven by several incidents that highlighted how chauvinism still permeates science. In April, evolutionary geneticist Fiona Ingleby of the University of Sussex in Brighton, UK, revealed on Twitter that *PLoS ONE* had rejected a paper that she wrote with a female colleague, after a reviewer said that adding “one or two” male co-authors would improve the analysis. The journal removed the reviewer from its database and asked the academic editor handling the paper to step down from its editorial board.

In June, Nobel-prizewinning biologist Tim Hunt drew widespread criticism when he spoke of his “trouble with girls” in laboratories. “You fall in love with them, they fall in love with you, and when you criticize them, they cry,” he said at an international science-journalism conference in Seoul. Hunt, who two days later resigned from his post as an honorary professor at University College London, said that he had meant to be light-hearted and that he had been “hung out to dry”, but the university did not reinstate him.

1,377

THE NUMBER OF PHYSICISTS WHO SHARED THE US\$3-MILLION BREAKTHROUGH PRIZE IN FUNDAMENTAL PHYSICS, WHICH WAS AWARDED IN NOVEMBER FOR RESEARCH ON NEUTRINOS.



US President Barack Obama announced the Precision Medicine Initiative in January.

October brought the biggest story of all: the revelation that renowned exoplanet hunter Geoffrey Marcy had sexually harassed multiple students over at least a decade. Marcy resigned from his post at the University of California, Berkeley, amid public outrage from colleagues at the university and in astronomy more widely (see page 464). The case has prompted soul-searching among scientific societies, and several are developing or re-evaluating policies intended to prevent sexual harassment at meetings and other events.

MOLECULAR FREEZE-FRAME Structural biologists uncovered unprecedented detail on life's molecular machinery this year, thanks to advances in a technique called cryo-electron microscopy (cryo-EM). Researchers can determine structures of cellular proteins by flash-freezing them, then photographing them at near-atomic resolution using an electron microscope. Cryo-EM has usurped X-ray crystallography in the past three years because it doesn't require proteins to be crystallized first, allowing researchers to analyse many more molecules.

Using the technique, biologists have mapped well over 100 molecular structures in detail this year, including the proteasome, which recycles damaged or unwanted proteins, and the spliceosome, which chops out pieces of messenger RNA before the sequence is translated into protein. This year also saw the sharpest cryo-EM structure so far — that of a bacterial enzyme involved in sugar breakdown — and researchers hope to bring this level of detail to medically important molecules.

MAKING MEDICINE PRECISE Tailoring treatments to individual patients has long been a goal in biomedicine, but US President Barack Obama gave this effort a big boost with his announcement in January of the Precision Medicine Initiative (PMI). As part of the US\$215-million programme, which will award its first grants next year, the NIH and partner organizations will recruit one million people across the country, collecting genetic information, health records and even data from electronic health-monitoring devices. Researchers will use the information to look for links between disease risk and genetic and environmental factors.

The PMI inspired other governments to bet on giant longitudinal studies of their own. Soon after Obama's speech, California announced a \$3-million initiative. And China is expected to launch its own large-scale project next year, which will take advantage of the country's considerable genomic-sequencing capacity.

Iceland showed this year what is possible with large numbers of human-genome sequences. In March, the Icelandic firm deCODE genetics in Reykjavik published four papers on its analysis of more than 2,600 full genomic sequences from Icelanders — the largest collection of human genomes from a single population. It described mutations linked to Alzheimer's disease and mutation rates in the Y chromosome. ■

Written by Monya Baker, Ewen Callaway, Davide Castelvecchi, Lauren Morello, Sara Reardon, Quirin Schiermeier & Alexandra Witze.

365 DAYS:
the year in science

IMAGES OF THE YEAR

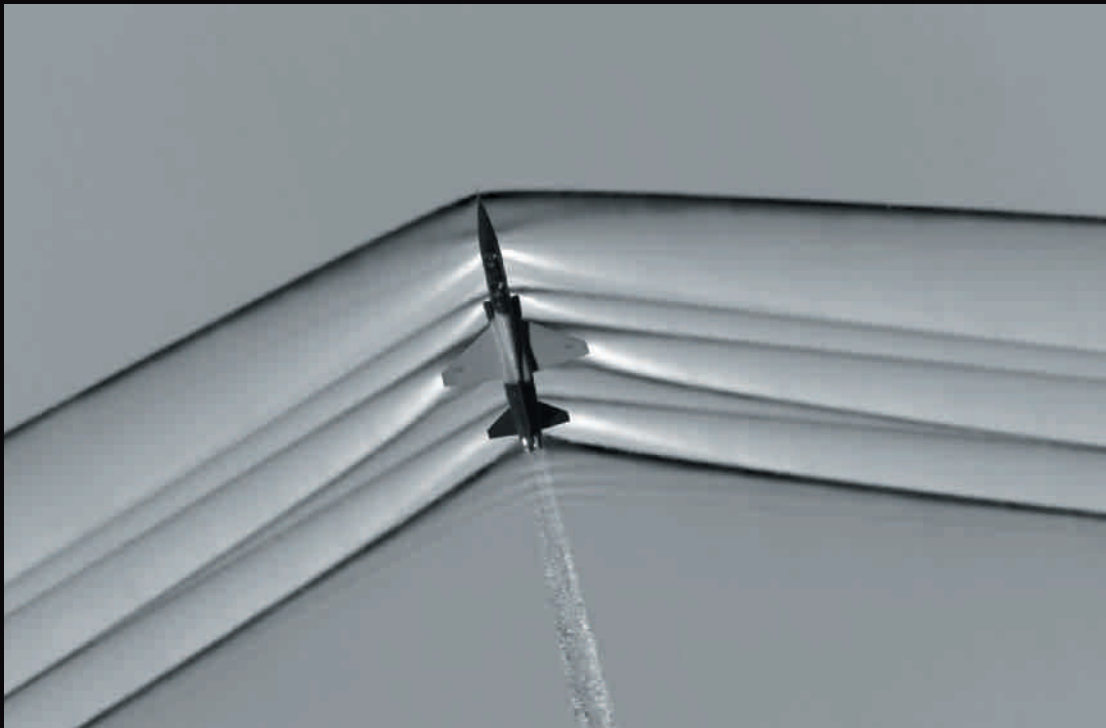
NASA's New Horizons probe won headlines and hearts this year as it sent back pictures of Pluto from the edges of the Solar System. But NASA scientists were not the only ones with images for us to wonder over. Animals at war, shock waves made visible and close-ups of objects normally beyond the limits of our vision were among the shots that caught the eye of *Nature's* art team.

Images selected by *Nature's* art and design team
Text by Daniel Cressey

REPTILE WARFARE

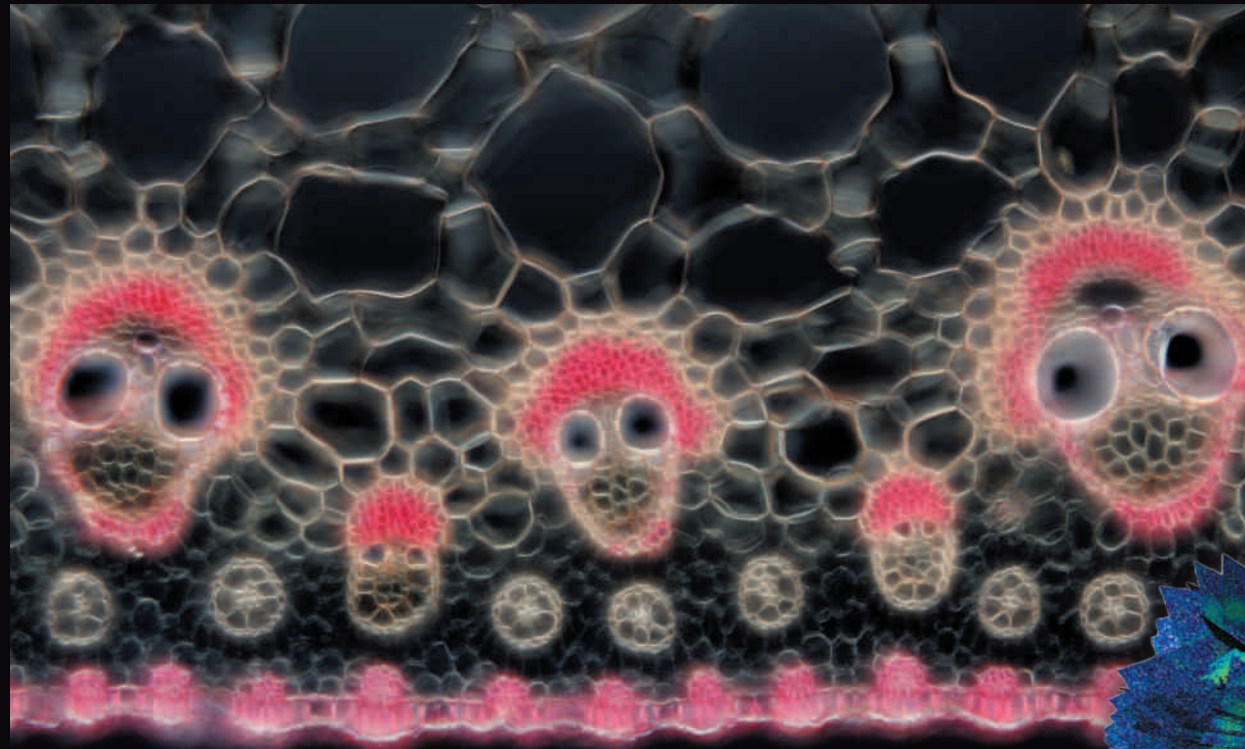
The biggest lizards on Earth — Komodo dragons — stage brutal fights over territory in Indonesia. This shot of such a bout was a finalist in the 2015 Wildlife Photographer of the Year competition.





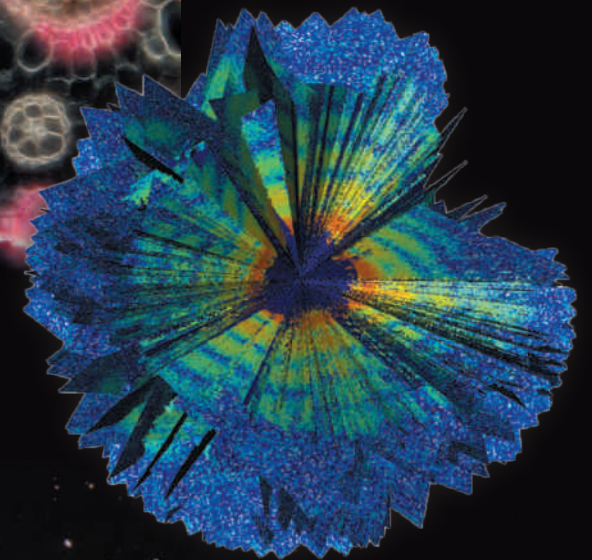
SUPERSONIC BOOM

The shock waves generated by a US jet moving at supersonic speed were imaged from another plane above the Mojave Desert. NASA researchers exploited a technique called schlieren photography, first developed in the nineteenth century by German physicist August Toepler, to capture changes in light as the jet passed through air of different densities.



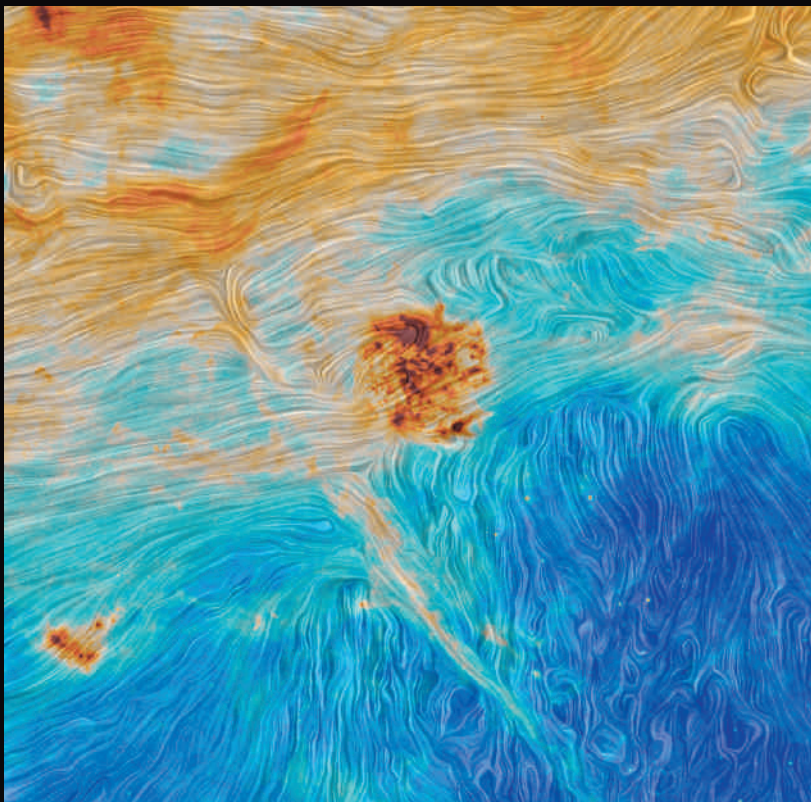
SPOOKY SLICE

These eerie, skull-shaped objects are actually a vital part of the papyrus plant (*Cyperus papyrus*). Photographed by David Maitland at 200 times life size, the image is a slice through the 'vascular bundles' that plants use to transport fluids through their tissues.



GOING VIRAL

It took hundreds of 2D snapshots of the large virus that infects *Acanthamoeba polyphaga* to produce this 3D structure. Researchers showed that powerful X-ray free-electron lasers could reconstruct a single particle of the giant virus despite its not being amenable to crystallization.

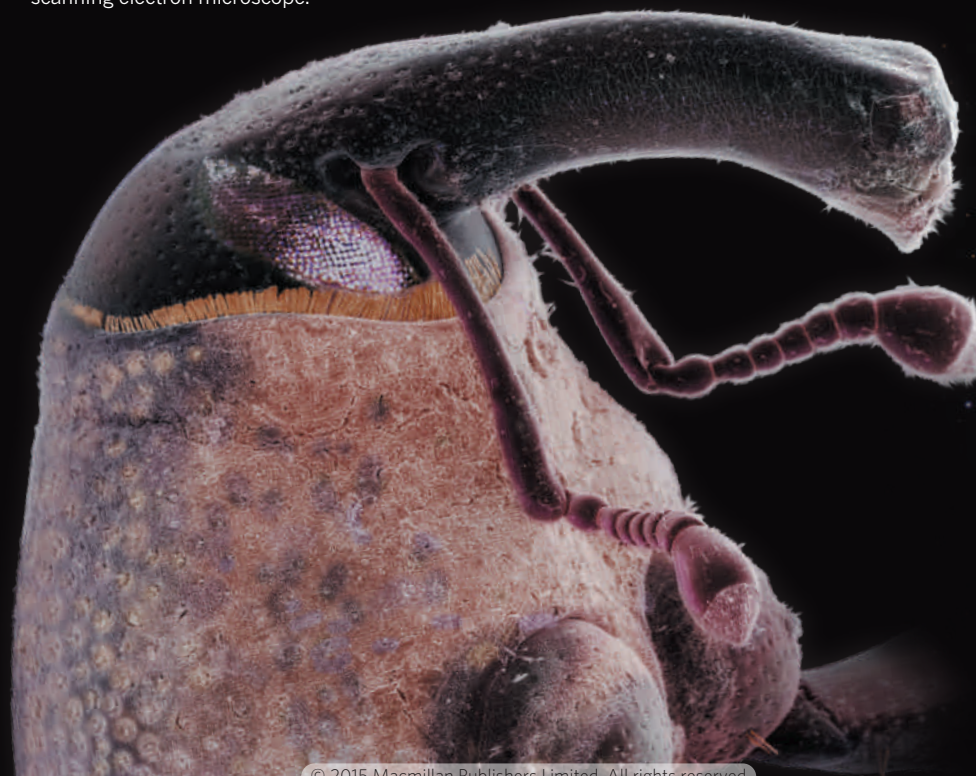


MAGELLANIC MAGIC

The Planck satellite provided a fresh view of the Large Magellanic Cloud (dark dots, centre) and the Small Magellanic Cloud (bottom left) — two galaxies close to our own Milky Way. The image uses data captured at microwave and sub-millimetre wavelengths.

THE WEEVIL'S HEAD

This detailed picture of the head of a boll weevil (*Anthonomus grandis*) was one of the winners in this year's Wellcome Image Awards. The head, which measures just millimetres across, was imaged using a scanning electron microscope.

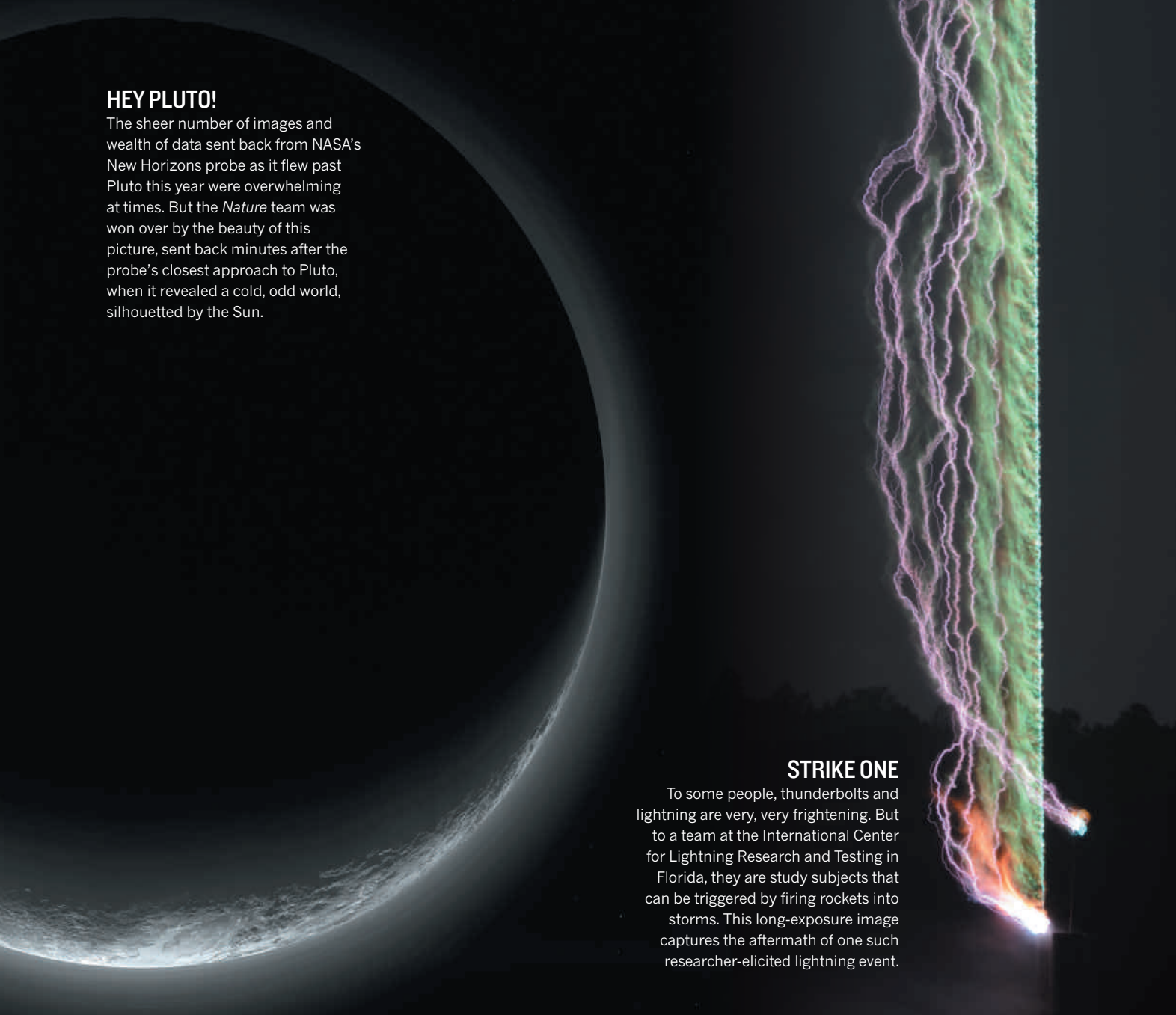


SPACE BUBBLE

This ghostly vision is a planetary nebula — the gently glowing remnants of a dying star. Nicknamed the Southern Owl Nebula, it was captured by the Very Large Telescope in Chile.

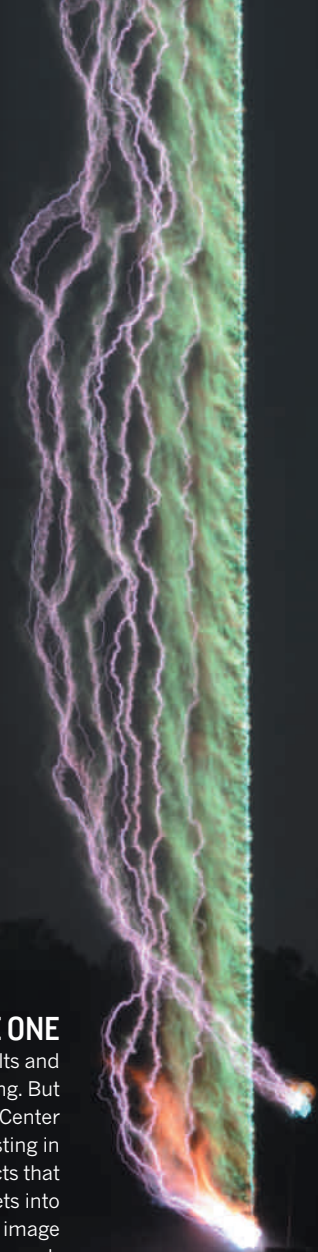
HEY PLUTO!

The sheer number of images and wealth of data sent back from NASA's New Horizons probe as it flew past Pluto this year were overwhelming at times. But the *Nature* team was won over by the beauty of this picture, sent back minutes after the probe's closest approach to Pluto, when it revealed a cold, odd world, silhouetted by the Sun.



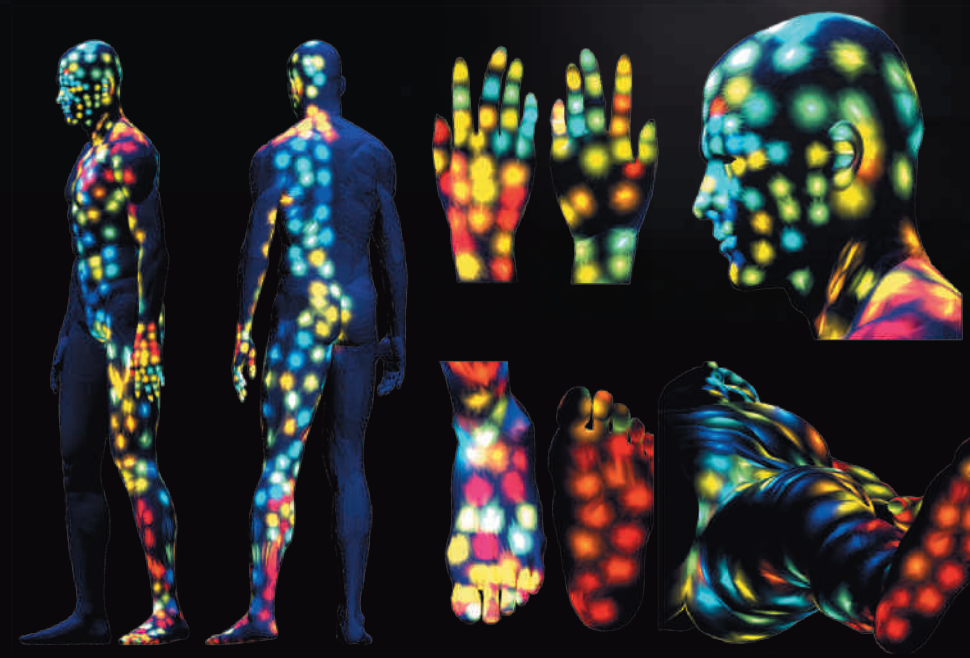
STRIKE ONE

To some people, thunderbolts and lightning are very, very frightening. But to a team at the International Center for Lightning Research and Testing in Florida, they are study subjects that can be triggered by firing rockets into storms. This long-exposure image captures the aftermath of one such researcher-elicited lightning event.



SKIN DEEP

This disco-map of the human body catalogues the chemicals and microbes found on the largest of all organs: the skin. Swabs from 400 sites on two healthy people were taken after the willing volunteers did not bathe for three days in the name of science.



KOMODO DRAGONS: ANDREY GUDKOV/WILDLIFE PHOTOGRAPHER OF THE YEAR 2015; SHOCKWAVES: NASA PHOTO; MEGALLANIC CLOUDS: ESA/PLANCK COLLABORATION; BOLL WEEVIL: DANIEL VARIKOW/ELLCOOME IMAGES; VASCULAR BUNDLES: DAVID MITLAND, COURTESY OF NIKON SMALL WORLD; VIRUS STRUCTURE: TOMAS EKEBERG/UPPSALA UNIV/AM. PHYS. SOC.; SPACE BUBBLE: ESO; PLUTO: NASA/JHUAPL/SWRI; LIGHTNING: UNIV. FLORIDA LIGHTNING RESEARCH GROUP; BODY SCAN: A. BOUSLIMAN ET AL. PROC. NATL. ACAD. SCI. USA 112, E2120-E2129 (2015); VULTURES: CHARLIE HAMILTON JAMES/WILDLIFE PHOTOGRAPHER OF THE YEAR 2015; DROUGHT: JUSTIN SULLIVAN/GETTY; MARS: NASA/JPL-CALTECH/UNIV. ARIZONA

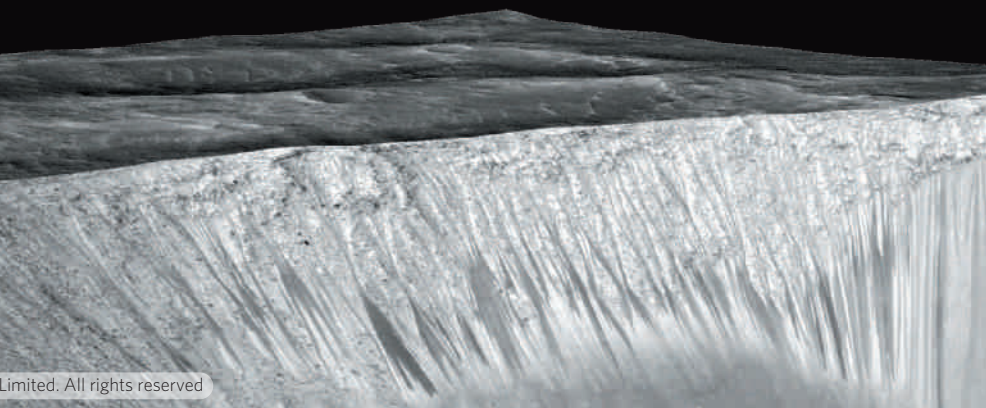


BODY OF EVIDENCE

Day-to-day life for African vultures is thrown into sharp focus by this 'carcass cam' shot. Although the scene is a bit gruesome, the birds' feeding habits play a key part in keeping the ecosystem healthy.

CALIFORNIA BURNING

The US 'golden state' has been hit hard by four years of severe drought. As locals and wildlife struggle to adapt to the dry spell, the frequency of fires, such as this one near Clearlake in August, has increased.



MARTIAN FLOWS

Planetary scientists have been finding water on Mars in different forms for some time now. But the dark streaks visible here are particularly exciting as they form part of the strongest evidence so far of liquid brine at the surface. The image was created by fitting images from NASA's High Resolution Imaging Science Experiment over a model of the terrain of the Garni Crater.

365 DAYS:
the year in science

FEATURE NEWS

Ten people
who mattered
this year.

NATURE'S 10

CHRISTIANA FIGUERES / JUNJIU HUANG / ALAN STERN / ZHENAN BAO
ALI AKBAR SALEHI / JOAN SCHMELZ / DAVID REICH / MIKHAIL EREMETS
CHRISTINA SMOLKE / BRIAN NOSEK

ILLUSTRATION BY PETER GROWTHER

CLIMATE GUARDIAN

*A dynamic leader
charted the path to a
new global climate
agreement.*

CHRISTIANA
FIGUERES

BY JEFF TOLLEFSON

Hours after the world's governments adopted a landmark climate accord this month, Christiana Figueres was all smiles on the dance floor of a boisterous night club in Paris. As the leader of the United Nations climate convention, she had spent five long years travelling the world to rally support among environmentalists, businesses and governments for the accord, in which 195 countries pledged to keep global warming to well below 2°C. But now here she was, leading conga lines and dancing to the Village People's classic 'Y.M.C.A.'

Asked whether she ever had any doubts, she flashed a smile, pulled her hands together as if in prayer and pointed skyward. "The stars are guiding us," she said.

Born into a politically powerful family in Costa Rica, Figueres came by her activism naturally. Her father led the republic's 1948 revolution and served as its first president. Her brother followed suit, with a term as president in the 1990s, and her mother served in the congress. Friends and colleagues credit Figueres for breaking out of her comfort zone in Costa Rica and jumping into the international environmental arena.

"In this country, being a Figueres means something," says Monica Araya, a former climate negotiator who founded Nivela, an environmental think tank based in Heredia, Costa Rica. "She built a whole career outside Costa Rica, and in a very important way she chose climate change as her activity."

Figueres attributes her environmental activism to the demise of a toad that disappeared from Costa Rica's Monteverde Cloud Forest Reserve. She saw one when she was young, but her daughters missed the chance. "That was a real awakening for me," she says, because rising temperatures have been linked to the toad's extinction. "I started reading into the topic, and before I knew it I was devoting my life to climate change."

In 1995, after stints in the Costa Rican government at home and abroad, Figueres created a non-profit organization in Washington DC to encourage Latin American engagement in the newly minted UN climate

convention. In parallel, she represented Costa Rica as a non-governmental climate negotiator — a move, Araya says, that helped to pave the way for other members of civil society to join the Costa Rican delegation. Over time, she became increasingly active in the governing secretariat of the UN convention and built up a reputation for getting things done. When Figueres was interviewed for her current post in 2010, she was asked what she would do if she were overruled by her boss. She offered up a quick joke: "Well, to begin with, I would fire him."

"She is brilliant, way above average, and she has a very well-developed sense of humour," says Marco Gonzalez, a friend and fellow Costa Rican who formerly headed the UN treaty organization that was built to phase out chemicals that damage the stratospheric ozone layer. "She brings success in her backpack."

Figueres took charge of an organization and a process that she describes as "in the garbage can" after the diplomatic meltdown at the Copenhagen climate conference in 2009. The secretariat had previously concerned itself mostly with national governments, but Figueres expanded its sphere by reaching out to local and regional governments as well as the business sector. "Her fingerprint is all over the intense presence of cities and businesses in Paris," says David Waskow, director of the International Climate Initiative at the World Resources Institute in Washington DC.

Figueres used all of her political skills to help herd governments towards the Paris agreement — and her roots in a developing country helped her to bridge the gulf between rich and poor nations, a division that had plagued past negotiations. Although current climate pledges fall short of the accord's ultimate goal, all nations have now committed to the battle against global warming.

Throughout the process, Figueres says she has been driven by the same sense of duty that spurred her father: the desire to protect and expand opportunities for those who are less fortunate. "I happened to choose a different battleground at the global level, but it's the same thing," she says. "We have a huge moral responsibility to do everything that we can to improve that situation." ■

EMBRYO EDITOR

A modest biologist sparked global debate with an experiment to edit the genes of human embryos.

BY DAVID CYRANOSKI

In April, Junjiu Huang published the world's first report of human embryos altered by gene editing. The news thrust rapid developments in gene-editing technology into the spotlight and ignited a huge debate about the ethical use of such tools. But Huang, a modest and soft-spoken molecular biologist at Sun Yat-sen University in Guangzhou, chose to stay out of the limelight.

Huang and his team used a powerful technique known as CRISPR-Cas9, which can be programmed to precisely alter DNA at specific sequences and has swept through biology labs in the past few years. He told *Nature* in April that he wanted to edit the genes of embryos because: "It can show genetic problems related to cancer or diabetes, and can be used to study gene function in embryonic development." In his study, he modified the gene responsible for the blood disorder β -thalassaemia.

Huang used spare embryos — from fertility clinics — that

could not progress to a live birth. And he expected his paper, which showed that the process created many unexpected mutations, to steer people away from the technology until it had been proved safe. "We wanted to show our data to the world so people know what really happened with this model," he said at the time. "We wanted to avoid ethical debate."

But the opposite happened: the ensuing discussion polarized the scientific community and nucleated several high-powered forums, including an international summit held in December in Washington DC. The general consensus is that gene editing is not yet ready for altering human embryos for reproductive purposes — and there are concerns that it could be adopted prematurely by rogue fertility clinics. Some scientists argue that the technique is permissible for research, whereas others say that this too should be forbidden for fear of a slippery slope.

Huang has been notably absent from the debate, and refused to be interviewed for this article. "Our paper was just basic research, which told people the risk of gene editing," he wrote in an e-mail. "It's like he's hiding," says Tetsuya Ishii, a bioethicist at Hokkaido University in Sapporo, Japan, who was at the US summit. "That's strange because there was nothing really ethically problematic about his research. He raised the issue, and that kind of drove discussions on the topic at the summit. That's a good thing." But Ishii says that Huang does "have some responsibility to address his critics", perhaps by discussing cases in which clinical use of gene editing could be worthwhile in the future.

Because of the risks, Huang predicted when his paper was published that it could take 50 or 100 years before the world saw a live-born, gene-edited baby. "But who knows, a decade ago, no one knew of CRISPR," he said. "We don't know what will happen." ■



COURTESY JUNJIU HUANG



ALAN STERN

PLUTO HUNTER

A single-minded planetary scientist brought the dwarf planet into focus.

BY ALEXANDRA WITZE

Alan Stern, planetary scientist and workaholic, doesn't sleep much at the best of times. In the days approaching 14 July — as the spacecraft he had dreamed about, worked for and slaved over for a quarter of a century neared its target — he was down to roughly three hours a night.

Stern, of the Southwest Research Institute in Boulder, Colorado, is the principal investigator for NASA's New Horizons mission, which in July became the first probe to visit Pluto. It whizzed just 12,504 kilometres above the dwarf planet's surface, in an extraordinarily choreographed fly-by that grabbed images, spectra and other scientific data — as well as headlines around the world.

Stern had been preparing for the day since 1989, when he and other young researchers hatched plans to visit the distant world. They submitted their proposal to NASA, and kept their hopes alive even when the agency killed plans for a Pluto mission in 2000 over budget concerns. After Congress revived funding for the concept, and NASA restarted the competition for proposals, Stern's team won with a lean design that would carry a few key instruments. "That meant a laser focus on getting it there," he says.

Stern is nothing if not laser focused. Under his leadership, New Horizons blasted off in January 2006 at a cost of US\$720 million, much less than earlier multibillion-dollar missions to the outer Solar System. His three children went through high school and into university with 14 July 2015 imprinted on their brains. When the day arrived, Stern

BILL INGALLS/NASA

MASTER OF MATERIALS

A chemical engineer is merging electronics with the human body.

BY ERIKA CHECK HAYDEN

Zhenan Bao rummages through a plastic box on her desk, eagerly pulling out samples of materials developed in her lab. She finds a thin, nearly weightless patch made of carbon nanotubes that attaches to the wrist like a sticking plaster and monitors the wearer's heart rate. Then she picks up an artificial skin that uses tiny carbon-nanotube sensors to detect touch; and a version of it that even features hair-like structures to more closely mimic real skin.

Bao, a chemical engineer at Stanford University in California and a founder of the field of thin, flexible organic electronics, shines a laser pointer through a sample of the nanotube material used in many of these devices. She laughs as the beam is diffracted into a spray of green dots on the wall, just as it would be when passing through a crystalline material. "That's how we know it has regular structure," she says.

Innovations in her field are often inspired by nature, she says: "If we can understand how to design materials with the same degree of complexity, we will be able to address real-world problems." A prime example is the creation of medical devices that can be worn or implanted to monitor blood sugar, send sensory signals and more.

Progress towards that goal has taken off this year, with Bao's lab among the leaders. In October, her team showed that its artificial skin could mimic the sense of touch (B. C. -L. Tee *et al. Science* **350**, 313–316; 2015). The researchers took inspiration from human skin, in which specialized nerves fire more rapidly as pressure increases, producing a code that the brain interprets as touch. Previous artificial touch sensors required power-hungry external devices to generate that code. But in Bao's sensors, pressure alters the oscillating frequency of microscopic circuits made from carbon nanotubes to generate the right kind of signals automatically.

Although Bao calls the final design "simple", it was a major accomplishment, says Polina Anikeeva, a neural-interfacing and materials scientist at the Massachusetts Institute of Technology in Cambridge. She notes that Bao has been working on perfecting these materials for years, and that her lab — which comprises around 40 chemists, chemical engineers and materials scientists — is highly interdisciplinary. "It's not just one idea," she says, "many ideas came together and made this possible."

"We have many years of work to do," says Bao, who hopes that the treasures she keeps in the plastic box will one day help to revolutionize health care. "But generally, the path is laid out." ■

COURTESY ZHENAN BAO

and the rest of Earth got to see Pluto up close for the first time. Among his favourite discoveries: ice mountains that tower as high as 4 kilometres, dune fields that may ripple across Pluto's surface, and skies that are tinted blue by atmospheric haze. A heart-shaped feature that showed up on images was a "public-relations bonanza", he says, inspiring people around the world to connect with the dwarf planet.

Stern's drive to explore new worlds is also reflected in his focus on public relations, says David Grinspoon, a researcher with the Planetary Science Institute in Tucson, Arizona, who is working with Stern on a book about the mission. Stern convened an eclectic group of artists, writers and visionaries in New York City months before the fly-by to pick their brains about ways to connect with the general public. "It wasn't your normal outreach team," Grinspoon says.

Stern pursues public engagement with a singular passion. He is known for seeking out — and scrutinizing — media coverage. Even during the most intense stages of the mission, Stern was tweeting prolifically and posting to Facebook while overseeing press releases.

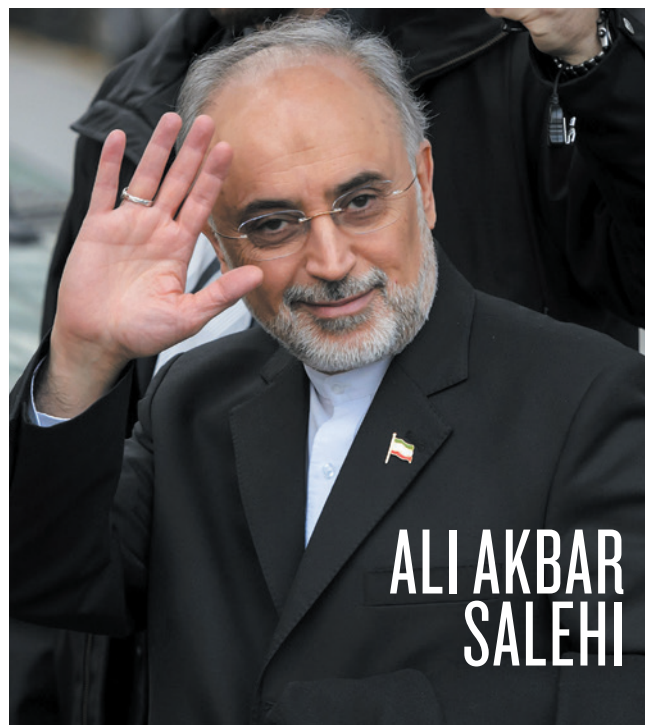
After the fly-by, Stern found himself swamped with speaking invitations. At an astronomy conference in Vermont, he talked for an hour, took questions for an hour and then met Pluto fans individually. Two university students told him that New Horizons was the best thing that had happened in their lifetime.

Months after the Pluto visit, some members of the team experienced a post-fly-by depression. Not Stern. He drives ahead as always, working on the data that will dribble back from the spacecraft until late 2016. He is also resuming work on the European Space Agency cometary mission Rosetta, on which he has an ultraviolet spectrometer instrument, and on plans to fly research payloads on suborbital spacecraft. He has a little more time for sleep these days, but not much.

And in October and November, New Horizons ignited its engines to set it on course to visit a second Kuiper belt object, this one on New Year's Day in 2019. If NASA approves the extended mission, Stern says, "I'm looking forward to finishing what we started". ■



ZHENAN
BAO



ALI AKBAR
SALEHI

NUCLEAR DIPLOMAT

The head of Iran's nuclear programme helped to forge a pact to keep it peaceful.

BY DAVIDE CASTELVECCHI

On 14 July 2015, Iran signed an agreement with six world powers to limit the country's nuclear development in exchange for lifted international-trade sanctions. If the deal is implemented successfully — still far from certain — it could ease years of tension over Iran's alleged efforts to build nuclear weapons and so allow the country to become a major player in global science. That an accord was reached at all, however, was due in no small measure to nuclear engineer Ali Akbar Salehi, who is head of the Atomic Energy Organization of Iran. He worked closely with his US counterpart, energy secretary Ernest Moniz, to iron out the deal's technical aspects.

Educated at the American University of Beirut and the Massachusetts Institute of Technology in Cambridge, Salehi returned to Iran after the Islamic revolution of 1979 and quickly rose to top posts in both academia and the government. By the 2000s, he had become the international face of Iran's nuclear programme — a man described as fiercely loyal to his country, but also a voice of reason to whom negotiators could appeal in times of crisis.

Salehi is said to be a deeply spiritual person who has the trust — and the ear — of the country's supreme leader, Ayatollah Ali Khamenei. And he is one of very few people to have held senior posts in both hardline and comparatively liberal governments.

This talent for building bridges is what enabled Salehi to work so effectively with Moniz during the negotiations, says Reza Mansouri, an astronomer at the Institute for Research in Fundamental Sciences in Tehran and a former deputy science minister of Iran; they shared the language of science. Mansouri, who has known Salehi for more than three decades, says that he has the modern, rational frame of mind that enables people to "agree on how to talk to each other". ■

FABRICE COFFRINI/AFP/GETTY



A VOICE FOR WOMEN

A senior astronomer worked to unmask a prominent sexual harasser.

BY ALEXANDRA WITZE

They came forward, one by one. Young female astronomers sought out Joan Schmelz and confided in her about the sexual harassment that they had endured. Schmelz, a solar physicist and chair of the American Astronomical Society's Committee on the Status of Women in Astronomy from 2009 to 2015, heard too many of these stories — and a lot of them involved the same man.

Schmelz told the women that they were not alone, and asked whether they wanted to talk to others who were in the same situation. Thanks in part to those introductions, four women eventually filed complaints. Their actions, which became public this year, led to the resignation of Geoff Marcy, a well-known exoplanet hunter at the University of California, Berkeley. It was one of the most dramatic episodes in a string of gender-equality controversies this year, including Nobel laureate Tim Hunt's dismissive comments about women working in the laboratory.

In astronomy, Schmelz's behind-the-scenes efforts to expose sexual harassment set the stage for a sea change in community understanding, says Meg Urry, an astronomer at Yale University in New Haven, Connecticut, and president of the astronomical society. After Marcy was outed, astronomy departments at universities and other institutions began frank discussions about unacceptable behaviour. "Without Joan,

ERIKA RODRIGUEZ FOR NATURE

GENOME ARCHAEOLOGIST

A big thinker helped to turn ancient genomics from niche pursuit to industrial process.

BY EWEN CALLAWAY

For most of its 30-year history, the field of ancient genetics has revolved around discovering exceedingly rare samples — a bone, a tooth — that harbour enough intact DNA to study. This year, population geneticist David Reich proved that it's possible to explore human history by powering through ancient genomes en masse.

Reich's genome factory has revealed mass migrations, the spread of farming and the roots of languages. Last month, his group at Harvard Medical School in Boston, Massachusetts, reported genome data from 230 people who lived in Europe and the Middle East over the past 8,000 years, tracking changes in skin colour, immunity and other traits (I. Mathieson *et al. Nature* <http://doi.org/9rb>; 2015).

At university, "I think I was sort of idealistic", Reich says. "I was

interested in grand unifying theories." For his first degree, he switched from sociology to physics. During his second, in biochemistry, he fell for human population genetics, and soon built a reputation for scientific rigour. In the late 2000s, plummeting sequencing costs and other advances made it easier to extract and analyse ancient DNA. Reich realized that by analysing the genomes of large numbers of people, he could see how immigration and interbreeding changed the genetics of entire regions.

In 2013, Reich opened his own lab devoted to sequencing ancient remains. Its scale was industrial from day one: the first human samples came from 66 individuals who had lived in what is now Russia, including members of a Bronze Age culture called the Yamnaya. In June, the team described a massive migration of Yamnaya people into Western Europe, some 5,000 years ago (W. Haak *et al. Nature* **522**, 207–211; 2015). It is not the only group powering through ancient genomes: the lab of Eske Willerslev at the Natural History Museum of Denmark in Copenhagen reached a similar conclusion (M. E. Allentoft *et al. Nature* **522**, 167–172; 2015).

Reich's team argued that the Yamnaya migration might also explain the radiation of Indo-European languages across Europe and Asia — advancing a problem that has vexed linguists for decades. By exploring the consequences of genetics for other fields, Reich "is trying to do something that a lot of geneticists might not", says David Anthony, an archaeologist at Hartwick College in Oneonta, New York. Reich is eager to see genetics inform other debates, such as those about the peopling of the Americas and the prehistory of India. "The invention of ancient DNA as a tool for studying the past is like the invention of a new scientific instrument, like a microscope," he says. "You can see into things that you couldn't see before." ■

STEPHANIE MITCHELL/HARVARD UNIV.

I don't think we would have seen this remarkable change," says Urry.

Women were comfortable sharing their stories with Schmelz because she had been through the same thing. Early in her career, Schmelz had found herself the target of harassment by her supervisor. "I was very isolated, and I didn't have anyone to confide in," she says. She only began to realize what had happened to her years later, in 1991, when attorney Anita Hill accused Clarence Thomas, a judge nominated for the US Supreme Court, of sexual harassment.

In 2011, Schmelz went public, through a blog post on the website of the Committee on the Status of Women in Astronomy. Then the Marcy stories started pouring in. "For a while I kept trying out how we could move forward — I contacted a lot of people, players in the community, to see if there was anything we could do for these women," she says.

Eventually the option emerged of filing complaints under the legislation known as Title IX, which prohibits sexual discrimination on campuses that receive federal funding. In July 2014, the first complaints hit Berkeley. "I wasn't sure it would ever happen," says Schmelz.

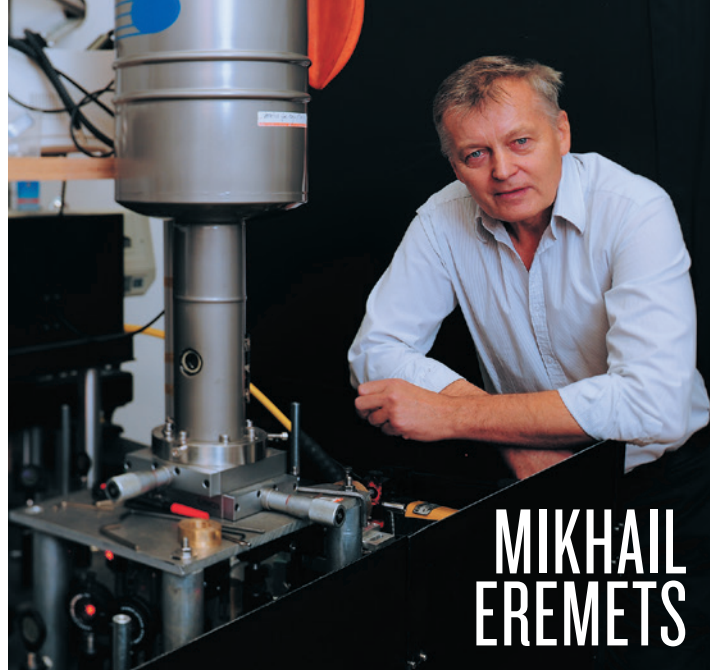
All this intense work took place as Schmelz led a busy career in solar astronomy. In June this year, she took a job as deputy director of Arecibo Observatory in Puerto Rico. Months later, the director resigned, leaving Schmelz in charge of the world's largest single-dish radio telescope.

She now lives just a block from the beach, which offers a much-needed respite when she can spare the time. But Schmelz knows that her work on harassment is not over. She would like to press universities to keep long-term records of complaints. In most institutions, there is no method for tracking whether there have been one, two or ten incidents reported against a given person over time.

"Let's find ways to take the pressure off the young women, so they can work on their science, write a thesis, without all of this extra added burden on them," says Schmelz. "Let's change the system." ■



DAVID
REICH



MIKHAIL
EREMETS

SUPER CONDUCTOR

Decades of diligence earned one physicist a record for resistance-free electricity.

BY EDWIN CARTLIDGE

As a young researcher during the 1970s and 1980s, Mikhail Erements proved to have a temperament well suited to life at the Institute for High Pressure Physics outside Moscow. The facilities were often abysmal, but the soft-spoken Belarusian was prepared to work around them — even dialling the same telephone number 100 times just to get a working line. "If I want to do something I am happy to repeat it many, many times," says Erements, who is now at the Max Planck Institute for Chemistry in Mainz, Germany.

That doggedness has served him well in his quest to understand how materials behave at pressures close to those of Earth's core — conditions that he recreates by squeezing tiny samples between the tips of two diamond 'anvils'. These experiments have been painstaking and repetitive, with results that never troubled the Nobel committee.

Until late 2014, that is, when Erements and his colleagues reported hints that pressurized hydrogen sulfide — the compound responsible for the smell of rotting eggs — can become a superconductor, allowing electricity to flow without resistance at a record-breaking 190 kelvin (-83°C) (A. P. Drozdov *et al.* Preprint at <http://arxiv.org/abs/1412.0460>; 2014). He and others published conclusive evidence — and measured an even higher temperature — in August (A. P. Drozdov *et al.* *Nature* **525**, 73–76; 2015). The advance has been hailed as a giant step towards the long-sought goal of room-temperature superconductivity and the promise of loss-free electrical transmission. It has certainly rocked the physics community, says Igor Mazin of the Naval Research Laboratory in Washington DC. Other materials have produced superconductivity at high temperatures, but the mechanism by which hydrogen sulfide operates has never achieved superconductivity above 40 kelvin.

No independent group has confirmed the result entirely, but Erements is already planning experiments to see whether hydrides doped with chemicals can superconduct at normal, atmospheric pressure — an essential step towards practical use. Having done most of his important work since turning 50, he feels he has plenty of research left in him. "In that sense I am still a young, growing scientist," he says. ■

CASRTEN COSTARD/MAX PLANCK INST. FOR CHEMISTRY



CHRISTINA SMOLKE

FERMENTING REVOLUTION

A synthetic biologist won a breakneck race to produce opioids in yeast.

BY ERIKA CHECK HAYDEN

Early this year, synthetic biologist Christina Smolke was in a dead-heat race with a handful of other labs to engineer a yeast strain capable of making opioids. These powerful pain-killing drugs are crucial in medicine, but they come solely from opium poppy crops that can have unpredictable yields. Scientists were seeking a more stable production method but faced a daunting hurdle: no one had been able to identify an enzyme that converts reticuline — a chemical building block of morphine and other narcotics — from one form to another.

Most other labs hunting for the enzyme were working to isolate it from poppies directly. But Smolke and her team at Stanford University, California, took a different approach: they combed through genetic databases, looking for snippets of sequence that looked as if they might be involved in reticuline metabolism. When they found a hit from several different poppy species, they ordered a synthetic version of the gene that had been built letter-by-letter by a machine. They plugged it into yeast and it worked. “I was super excited, really proud and also relieved,” Smolke says. “It was a bit of a Hail Mary.”

The discovery enabled Smolke’s lab to stitch together a pathway of 23 different genes from plants, mammals, bacteria and yeast to produce the world’s first narcotic through synthetic biology (S. Galanie *et al. Science* **349**, 1095–1100; 2015). It was a crowning achievement for a

ROD SEARCEY/STANFORD ENGINEERING

BIAS BLASTER

A psychologist pledged to improve reproducibility in science.

BY BRENDAN MAHER

When Brian Nosek was a graduate student in experimental psychology, he started working on the implicit-association test, which reveals people’s unconscious prejudices with the push of a button. Tap right every time a male name appears on a screen, for example, and left for a female name. That’s easy — but add some stereotypically male or female roles into the mix and things get interesting. Even the most liberal minds will sometimes stall when asked to press the same button for the word ‘executive’ and for the name ‘Susan’.

The tests are challenging, informative and kind of fun. So in 1998

Nosek convinced his mentors, who had developed the test, to put it online. It was a success: about a million people per year now take the test for research, corporate training and other reasons. “It really spread the word about what unconscious bias is,” says Betsy Levy Paluck, a social psychologist at Princeton University, New Jersey.

For Nosek, a key demographic still needs to be educated about their biases: scientists. Nosek is convinced that researchers are unconsciously influenced by their hypotheses, that these biases can be seen in common practices that distort the interpretation of data such as *p*-value hacking, and that they are major drivers of the much-discussed crisis in research reproducibility. In 2013, Nosek took leave from his post at the University of Virginia in Charlottesville to co-found the Center for Open Science (COS), a non-profit company that builds tools to facilitate better research methodology. It hit several milestones this year, accumulating US\$18 million in funding and a staff of 68. Nosek also co-authored a set of guidelines for transparency and openness that more than 500 journals have signed up to (B. A. Nosek *et al. Science* **348**, 1422–1425; 2015).

But the COS’s most visible output in 2015 was the Reproducibility Project, an ambitious attempt to re-test seminal findings in

UNIV. VIRGINIA

Ones to watch

2016

FABIOLA GIANOTTI

DIRECTOR-GENERAL OF CERN

Gianotti will take charge at the European lab as its Large Hadron Collider clocks up record high-energy particle collisions — and as hopes of the next big discovery soar.

GABRIELA GONZÁLEZ

SPOKESPERSON AT ADVANCED LIGO

If rumours that this observatory has detected gravitational waves prove true, one of the most elusive predictions of the general theory of relativity would be confirmed.

KATHY NIAKAN

STEM-CELL BIOLOGIST, FRANCIS CRICK INSTITUTE

By applying for approval to edit the genomes of human embryos, Niakan has placed herself at the front of the fast-moving, controversial CRISPR-Cas9 field.

DEMIS HASSABIS

CO-FOUNDER, DEEPMIND

There is intense curiosity about what will emerge next from Hassabis's efforts to combine neuroscience and machine learning at the Google-owned firm.

YANG WEI

HEAD, NATIONAL NATURAL SCIENCE FOUNDATION OF CHINA

Yang will be influential at this growing basic-research agency as China overhauls its funding systems and sets its next 5-year plan.

biological wunderkind who started her own lab at the California Institute of Technology in Pasadena at the age of just 28. The opioid-producing yeast cells contain the most complex synthetic-biology pathway developed so far, and mark a turning point for the field by showing how step-by-step engineering can turn microbes into drug factories. "This will significantly impact our future ability to produce many more chemicals through biotechnology," says Jens Nielsen, a synthetic biologist at the Chalmers University of Technology in Gothenburg, Sweden.

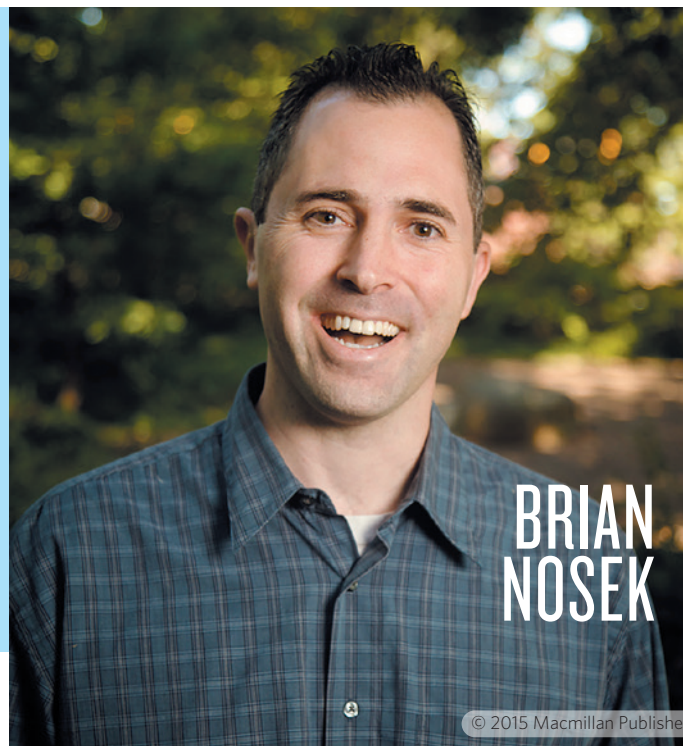
Much of the news coverage of the work, however, stirred fears about how it could foster new ways to easily manufacture illegal drugs — and some scientists have argued for tighter regulation of the growing field. Smolke counters that existing regulations already restrict the production and distribution of narcotics; any lab that wishes to work with the yeast strain reported in her paper, for instance, must be licensed by the US Drug Enforcement Administration. So far, no one has requested the strain.

In a bid to ground the debate in reality, Smolke, her husband — fellow Stanford synthetic biologist Drew Endy — and another colleague this year attempted to brew opioids using her lab's strain and standard beer-making equipment (D. Endy *et al.* Preprint at bioRxiv <http://doi.org/9t2>; 2015). The set-up produced only a trace amount of reticuline and none of the

downstream chemical, thebaine, that is used to synthesize commercial drugs such as oxycodone and oxymorphone — suggesting that it would be difficult for the average home-brewer to start making these pharmaceuticals. (The scientists' positive fermentation control, an English ale, was "palatable", the manuscript notes.)

Smolke co-founded a company, Antheia, based in Palo Alto, to produce opiate drugs in yeast commercially, and specialists in the field suspect that more will follow. But some onlookers are circumspect. Plant biologist Ian Graham at the University of York, UK, says that it will be hard to beat poppies. "Where plants already do it very well, the arguments for taking a synthetic-biology route are much less convincing," he says.

For Smolke, the goal is not merely to copy plants, but to engineer opioids that are free of side effects such as dependency and addiction. Sitting in the office of a Palo Alto incubator space, wearing jeans and grey Converse sneakers to a meeting with the co-founders of Antheia, Smolke can appear casual — but the intensity that has propelled her to the pinnacle of her field is tangible. For her, the year's accomplishments are just part of a quest to understand and improve on opioids, which are among the most complex natural chemicals. "It's a very powerful approach to take inspiration from nature and go beyond it," she says. ■



BRIAN NOSEK

100 psychology papers (Open Science Collaboration *Science* <http://doi.org/68c>; 2015). The decision to run the project "was quite brave of him", says Dorothy Bishop, a neuropsychologist at the University of Oxford, UK, because poor results could tarnish the field's reputation. In the end, 61 of the findings could not be replicated — but the outcome was mostly received well, something for which many psychologists credit Nosek's careful diplomacy and can-do approach.

Nosek is pushing researchers to adopt practices that will improve reproducibility, including preregistering studies, tracking the results in an open way and publishing them whether they are positive or negative. It will be a dramatic culture change, says Bishop, who has begun using systems developed by the COS for her own research. "Yes, it creates a lot more work. You have to document and check it very thoroughly. But it's not a bad thing to be slowed down a bit."

A second reproducibility project that is focused on findings in cancer biology should begin releasing results next year, and Nosek says that negotiations are in the works for similar projects in ecology and computer science. No one operates completely free of bias, he says, and that includes him. "I try to have some humility and understanding that I am as prone to these behaviours as anyone else." ■

COMMENT

EQUALITY Family policies are necessary, but not sufficient **p.471**

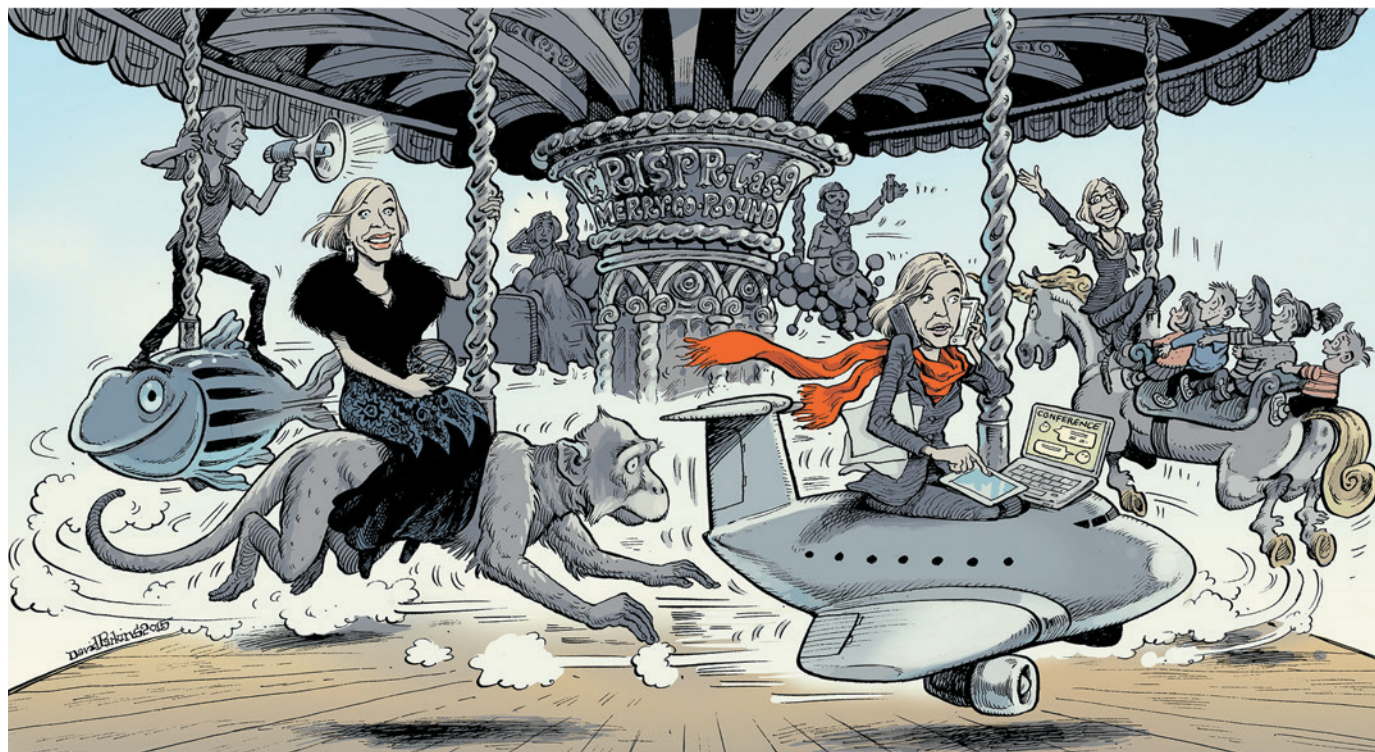


ECOLOGY George Schaller, pioneering field biologist, in conversation **p.474**

CULTURE On the beautiful relationship between mathematics and art **p.476**

OBITUARY Maurice Strong, architect of UNEP, remembered **p.480**

ILLUSTRATION BY DAVID PARKINS



My whirlwind year with CRISPR

Jennifer Doudna, a pioneer of the revolutionary genome-editing technology, reflects on how 2015 became the most intense year of her career — and what she's learnt.

Some 20 months ago, I started having trouble sleeping. It had been almost two years since my colleagues and I had published a paper¹ describing how a bacterial system called CRISPR–Cas9 could be used to engineer genomes (see ‘Based on bacteria’).

I had been astounded at how quickly labs around the world had adopted the technology for applications across biology, from modifying plants to altering butterfly-wing patterns to fine-tuning rat models of human disease. At the same time, I'd avoided thinking too much about the philosophical and

ethical ramifications of widely accessible tools for altering genomes.

Questions about whether genome editing should ever be used for non-medical enhancement, for example, seemed mired in subjectivity — a long way from the evidence-based work I am comfortable with. I told myself that bioethicists were better positioned to take the lead on such issues. Like everyone else, I wanted to get on with the science made possible by the technology.

Yet as the uses of CRISPR–Cas9 to manipulate cells and organisms continued to mount, it seemed inevitable that researchers

somewhere would test the technique in human eggs, sperm or embryos, with a view to creating heritable alterations in people. By the spring of 2014, I was regularly lying awake at night wondering whether I could justifiably stay out of an ethical storm that was brewing around a technology I had helped to create.

GROWING EXCITEMENT

“I hope you're sitting down because it's unbelievable how well it's working.” That was the verdict, delivered in December 2012, of a colleague who had been experimenting ►

► with CRISPR–Cas9. It reflected my own lab's experience, and that of others who had contacted me that autumn to share their excitement about the genome-editing technology.

It often takes years for a new molecular tool to take hold. Yet even before the end of 2012 — just a few months after my colleagues and I had published our initial study — at least six papers describing different uses of CRISPR–Cas9 for genome engineering had been submitted for publication.

In early 2013, several papers, including some describing how the technology could be used to edit the genomes of human stem cells and to alter a whole organism (the zebrafish), were an early indication of the coming tsunami^{2,3}. By the end of 2014, scientists had — among other things — used CRISPR–Cas9 to enhance pest resistance in wheat, reproduce the carcinogenic effects of specific chromosome translocations in mouse lungs and correct a mutation in adult mice that in humans causes the disease hereditary tyrosinaemia^{4–6}.

An ethically more complicated potential use of CRISPR–Cas9 was underscored in February 2014, when researchers described how they had used it to make precise changes to the genomes of cynomolgus monkey embryos⁷. (Cynomolgus monkeys are so genetically close to humans that they are commonly used to model human genetic disease.) The monkeys that developed — through implantation of the embryos into surrogate mothers — carried the genetic changes in most of their cells, including their eggs or sperm. This meant that the alterations could be passed down to future generations.

I was alerted to the paper by reporters seeking my comments on the research. After reading the preprint, I gazed out of my office window and across the San Francisco Bay and pondered how I would feel if the next reporter to contact me wanted to know about genome-editing work involving human embryos. “How long will it be before someone tries this in humans?” I wondered aloud to my husband over breakfast the next day.

At the same time, I had been receiving e-mails from people facing potentially devastating genetic predicaments. In one message, a 26-year-old woman told how she had discovered that she carried the *BRCA1* mutation, which gave her a roughly 60% chance of developing breast cancer by the time she was 70. She was considering having her breasts and ovaries removed, and wanted to know whether the approaches made possible by CRISPR–Cas9 meant that she should hold off.

The monkey study and interactions with patients or their relatives weighed on me. Every day brought a new influx of papers describing research using CRISPR–Cas9. My inbox was full of requests from researchers seeking advice or collaboration. All

BASED ON BACTERIA

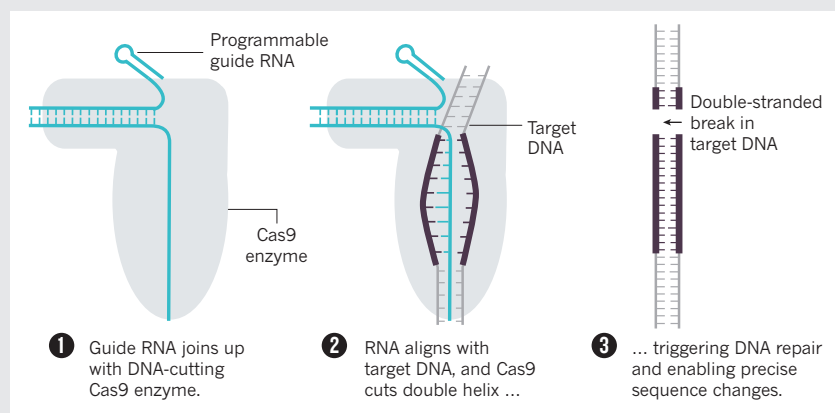
How CRISPR–Cas9 works

Clustered regularly interspaced short palindromic repeats, or CRISPRs, are repeating sequences found in the genetic code of bacteria. They are interspersed with ‘spacers’ — unique stretches of DNA that the bacteria grab from invading viruses, creating a genetic record of their malicious encounters.

On a repeat encounter with a virus, a bacterium can produce a stretch of RNA that matches the viral sequence, using the material in its spacer archive. This

‘guide RNA’ teams up with DNA-cutting Cas enzymes, encoded by nearby CRISPR-associated genes, to seek out and ‘cleave’ the matching viral sequences, stopping the virus from replicating.

By engineering the guide RNA, researchers can programme Cas enzymes — most commonly Cas9 — to match the DNA at specific sites that they want to cut in a cell's genome. This triggers a DNA repair that can result in precise sequence changes to the gene of interest.



this activity could have a direct impact on human life, yet most people I knew outside of work — neighbours, extended family members, parents of my son's classmates — remained largely oblivious. I felt as though I was living in two separate worlds.

Towards the end of 2014, my unease outweighed my reluctance to step into a more public discussion. It was clear that governments, regulators and others were unaware of the breakneck pace of genome-editing research. Who besides the scientists using the technique would be able to lead an open conversation about its repercussions?

THE ETHICS DEBATE

My first serious foray into the ethics was a one-day conference in January in California's Napa Valley, which I helped to organize and which was sponsored by the Innovative Genomics Initiative. Eighteen of us (scientists, bioethicists, a film-maker and an administrator from the University of California, Berkeley) discussed how genome engineering could affect health care, agriculture and the environment. In particular, we talked about issues

surrounding the modification of the human germ line — eggs, sperm and embryos.

Shortly after the meeting, we published a perspective article in *Science*⁸ that urged the global scientific community to refrain from using any genome-editing tools to modify human embryos for clinical applications at this time. We also recommended that public meetings be convened to educate non-scientists and to enable further discussion about how research and applications of genome engineering might be pursued responsibly.

Since the Napa meeting, I have given more than 60 talks about CRISPR–Cas9 — at schools, universities and companies, and at some two dozen conferences across the United States, Europe and Asia. I have spoken about it before the US Congress; talked to staff members at the White House Office of Science and Technology Policy, which provides science advice to the US president; and answered questions from the governor of California, among many others. These discussions have pushed me far outside my scientific comfort zone.

I am a biochemist; I haven't worked with animals, human subjects or human tissues, and there was a lot that I didn't know about the ethical difficulties inherent in other areas of research such as cloning, stem cells



and *in vitro* fertilization. I have relied on the generosity of colleagues who have helped to educate me — about how experiments involving human subjects or tissues are regulated in different countries, for example, and how ethical difficulties stemming from *in vitro* fertilization have been handled historically.

This year has been intense — and intensely fascinating. At times I have wished that I could step off the merry-go-round, just for a few minutes, to process everything. Ensuring that my travel and other commitments do not disrupt the progress of my lab members has been a priority, but working with them has increasingly involved meeting at night or on weekends, or conferring by e-mail or Skype. For now, time for my beloved vegetable garden and for hikes into the wilds of California with my 13-year-old son is gone.

Almost three years after a colleague warned me that a “tidal wave” of research, discussion and debate involving CRISPR–Cas9 was coming, I still don’t know when the wave will crest. But as the year ends, there are some things of which I am sure.

BROADENING THE CONVERSATION

With only 18 attendees — all from the United States and most of whom were scientists — the Napa meeting could only ever be a starting point for a broader conversation. But the meeting, and the commentary that resulted, were important on two fronts.

By mid-2014, I was concerned that CRISPR–Cas9 would be used in a way that was either dangerous, or perceived to be dangerous, before scientists had communicated enough about it to the wider world. I wouldn’t have blamed my neighbours or friends for saying, “All this was going on and you didn’t

tell us about it?” The *Science* perspective, and a related Comment published in *Nature* the week before⁹, helped to convey the message that those leading the work recognized that they had a responsibility to voice concerns.

The discussion initiated by these articles — which grew more urgent when a study was published in April in which CRISPR–Cas9 was used to modify the genomes of non-viable human embryos¹⁰ — also helped to set in motion the multitude of hearings and summits that have happened around the world since. The most prominent of these occurred in Washington DC earlier this month when the Chinese, US and UK science academies co-hosted a meeting on gene editing in humans.

With science now so influenced by international collaboration, scientists can in principle shape the direction of the global scientific enterprise to some extent through self-censorship. It seems obvious to me now that engendering more trust in science is best achieved by encouraging the people involved in the genesis of a technology to actively participate in discussions about its uses. This is especially important in a world where science is global, where materials and reagents are distributed by central suppliers and where it is easier than ever to access published data.

I am excited about the potential for genome engineering to have a positive impact on human life, and on our basic understanding of biological systems. Colleagues continue to e-mail me regularly about their work using CRISPR–Cas9 in different

“These discussions have pushed me far outside my scientific comfort zone.”

organisms — whether they are trying to create pest-resistant lettuce, fungal strains that have reduced pathogenicity or all sorts of human cell modifications that could one day eliminate diseases such as muscular dystrophy, cystic fibrosis or sickle-cell anaemia.

But I also think that today’s scientists could be better prepared to think about and shape the societal, ethical and ecological consequences of their work. Providing biology students with some training about how to discuss science with non-scientists — an education that I have never formally been given — could be transformative. At the very least, it would make future researchers feel better equipped for the task. Knowing how to craft a compelling ‘elevator pitch’ to describe a study’s aims or how to gauge the motives of reporters and ensure that they convey accurate information in a news story could prove enormously valuable at some unexpected point in every researcher’s life. ■ [SEE NEWS REVIEW P.449](#)

Jennifer Doudna is a Howard Hughes Medical Institute investigator and professor of molecular and cell biology, and of chemistry, at the University of California, Berkeley, Berkeley, California, USA. e-mail: doudna@berkeley.edu

1. Jinek, M. *et al. Science* **337**, 816–821 (2012).
2. Hsu, P. D., Lander, E. S. & Zhang, F. *Cell* **157**, 1262–1278 (2014).
3. Doudna, J. A. & Charpentier, E. *Science* **346**, 1258096 (2014).
4. Wang, Y. *et al. Nature Biotechnol.* **32**, 947–951 (2014).
5. Maddalo, D. *et al. Nature* **516**, 423–427 (2014).
6. Yin, H. *et al. Nature Biotechnol.* **32**, 551–553 (2014).
7. Niu, Y. *et al. Cell* **156**, 836–843 (2014).
8. Baltimore, D. *et al. Science* **348**, 36–38 (2015).
9. Lanphier, E., Urnov, F., Haecker, S. E., Werner, M. & Smolenski, J. *Nature* **519**, 410–411 (2015).
10. Liang, P. *et al. Protein Cell* **6**, 363–372 (2015).

Scientists must work harder on equality

Astronomer **Meg Urry** reflects on a turbulent year for women in science.

Gender equality in science made headlines repeatedly this year. Nobel-prizewinning biochemist Tim Hunt made his ill-advised quip about women in labs; Shrinivas Kulkarni, an astrophysicist at the California Institute of Technology, called astronomers and their telescopes “boys with toys”; and in a much more serious matter, astronomer Geoff Marcy resigned from his post at the University of California, Berkeley, after public

disclosure that he had sexually harassed female students. More quietly, there were rumours that at least three astronomers had been dismissed, and in some cases scrubbed from institutional websites.

None of these incidents were in any way related to motherhood, which was — and is — too often invoked to explain the dearth of women in science. (Gender is of course neither binary nor necessarily stationary; that I talk about ‘women’ and ‘men’ in this piece is

not meant to obscure that point.)

As the mother of two amazing women, I would say that family issues are the least of the problem. It is unquestionably true that employers must improve support of families, with progressive policies on paid parental leave, care of the elderly, high-quality on-site child care, and tenure ‘clock stops’.

But if inequality were all about family issues, why has women’s participation in the life sciences grown so much faster over the

past three decades than in physics or engineering? (see 'Running the gauntlet'). Why, in the United States, where I have worked in the scientific enterprise for nearly four decades, does astronomy have twice the percentage of women that physics does, despite requiring a very similar skill set? And if fixing the disproportionate burden of family care on women is all that matters, countries that have strong family-support systems — such as Sweden and Denmark — would have greater participation of women in science than in the United States, which languishes near the bottom of parental pay and leave leagues.

It has been shown that women without children generally do not advance any faster or further than women with families. In their ground-breaking 2002 paper¹, 'Do Babies Matter', researchers Mary Ann Mason and Marc Goulden showed that women with children who remain in full-time academia are no worse off than women without children. Both groups lag well behind men — especially men with children, who lead everyone else.

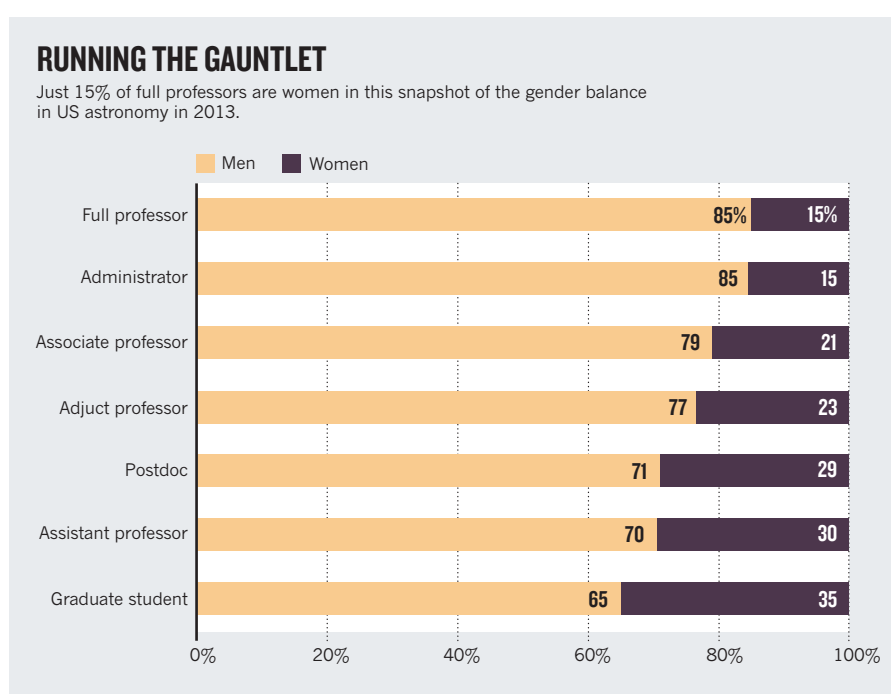
Clearly, strong family-support policies are not the whole story.

CHAMPIONS AND CRITICS

Every major criterion on which scientists are evaluated, for hiring, promotion, talk invitations or prizes, has been shown to be biased in favour of (white) men. These include authorship credit², paper citations³, funding⁴, recruitment⁵, mentoring and tenure. For example, although women publish fewer papers than men, there is some evidence that on average they are longer and more complete, and that this difference vanishes if one corrects for funding level and research-group size.

Women in male-dominated careers face obstacles that are often invisible and usually unacknowledged (just read Virginia Valian's 1998 book, *Why So Slow? The Advancement of Women* (MIT Press) and the papers described in her annotated bibliography). I have experienced many of these obstacles. People often have a just a little more certainty that the man is a genius and a little more doubt that the woman will make the grade. Her contribution to the paper — was it her student's brilliance or her husband's work? Her work is risky and unlikely to succeed whereas his is revolutionary; hers is pedestrian while his is reliable. Men have champions; women have critics.

Letters of recommendation for women are shorter than letters for men. They are less detailed and are filled with 'grindstone' adjectives (such as 'hard-working', 'determined' and 'dependable') rather than superlatives ('brilliant', 'creative', 'outstanding'). They are more likely to mention personal characteristics ('likeable', 'friendly', 'helpful') and more likely to mention gender and parenting issues (for instance, "she did all this



work while having two children"). These differences hold true whether the writer is male or female⁶. Women are invited to give fewer talks and asked to sit on fewer scientific organizing committees and prestigious committees — yet they do much of the everyday committee work.

As a senior female astrophysicist, my proposals to use the Hubble Space Telescope — equivalent to winning funding of US\$100,000 if granted — are less likely to succeed than those of my male colleagues (or my junior female colleagues)⁷. The difference is not statistically significant in any one review cycle, but after 25 years, it is clear that senior women are systematically less successful than their male counterparts, at a level of a few per cent per cycle. This is striking because almost all Hubble proposals are written by large teams that include both men and women, so the quality of the text does not depend on the gender of the principal investigator.

I am less likely to be nominated for a prize or honour⁸. I am more likely to be paid less (and was, for many years). In my experience, women are more likely to report having received gratuitously rude referee reports on their papers. (Whether the criticism is nastier or the sting is felt more acutely is not clear.)

Meanwhile, in my experience, women spend much more time teaching, mentoring and doing outreach than do their male colleagues. And this work is often not valued. One woman I know was described as having succeeded in her research "despite all the time she spent on outreach", as though her choice to attract girls to science was misguided. I would have described her as a superstar, who accomplished a very difficult

(what some might call a 'highly risky') scientific measurement while creating an innovative new course and investing time in the future of her discipline.

And we wonder why the attrition of women remains greater than attrition of men at every level in the scientific hierarchy.

TIME FOR CHANGE

We should not forget that within living memory many Western democracies overtly — not just covertly — discriminated against women. Before 1969, some of the best US research universities did not admit women as undergraduates (two being Yale University in New Haven, Connecticut, and Johns Hopkins University in Baltimore, Maryland, where I was educated). Equal-pay acts were not passed until 1963 in the United States, 1970 in the United Kingdom, and decades later in other parts of Europe. As recently as 1990, there remained elements of voting that were open only to men in one part of Switzerland. When I first applied for assistant-professor positions about 25 years ago, some universities still had anti-nepotism rules, which were a real problem for scientific couples.

But gender inequality today is not about discrimination in the past. In the United States, institutes established since the 1980s are just as biased as the oldest in the land. California's Silicon Valley, which has flourished in the past few decades, has an abysmally low number of women in business leadership positions.

I have heard colleagues say, "women don't want faculty jobs — the work is too hard, it's incompatible with having a life". Apart from this being nonsense, the answer is not to ignore half of the brains. Rather, it is to create

a more humane workplace, in which impact and quality of work have greater weight than monastic devotion and 100-hour work weeks.

What prompts people to conclude that women don't want faculty jobs? It is, in part, because the presence of women in the applicant pool for such jobs can be much lower than the fraction of women who are qualified for the positions — simply because men apply to many more jobs, on average, than women do. The low fraction of women has nothing to do with lack of interest.

Social-science research on confidence hints at why this might be the case (see, for example, ref. 9). Women tend to apply only to jobs for which they feel they have a fighting chance, either because the qualifications listed in the job advertisement match theirs or because the institution is one that they think they are good enough to join; men apply regardless. Recruiters should note that female applicants, being more selective in their attempts, are likely to be well suited to the position that they have applied for.

When I give a colloquium at a university whose physics department lacks female faculty members, I often ask: "Have you thought about hiring women?" The answer is usually earnest: "Oh yes, we definitely want to do that, but we want to hire the best." Do my hosts realize how insulting it is to imply those two goals are mutually exclusive?

Recently, a colleague worried openly about young men who, in the face of added competition from women, might not land that coveted assistant-professor position. If a woman of equal ability were hired affirmatively in place of a man, he suggested, the unsuccessful male applicant should be compensated with \$100,000. My jaw dropped. By that reasoning, shouldn't we compensate the thousands of women or other underrepresented scientists who were preferentially not hired over the past 50 years, despite being as talented as — or substantially more so than — the men who got the jobs?

Rather than focusing on what young men are losing when they have to compete with talented women, we should be asking what research is losing by not having the full participation of women. Sometimes, science feels increasingly homogenous, with professors training graduate students to think like them, and sameness being valued.

As I (and many others) have pointed out several times, the failure to hire women and minorities in science is a guarantee that the best are not being hired. The old canard that there aren't any women or there aren't any people of colour does not hold up. When you

look, they are there. And they bring talent that we desperately need, not to mention huge value as role models for students, who are so much more diverse a group than the faculty.

BEST PRACTICE

Many practical steps increase the likelihood of hiring and retaining women and other underrepresented scientists. For example, before evaluating applicants for a position, a search committee should agree on the set of desired qualities (subfield of research, teaching ability, publication record, contribution to diversity, ideas for student projects, research funding, and so on). When each candidate is evaluated in those categories, bias in the outcome is reduced.

Institutions can tone down elitist language in job advertisements without hurting their programme — status depends on quality, not adjectives. Women can be more likely to apply to institutions that describe themselves as 'collegial' and 'student-oriented' than 'top-rated' and 'world-class'.

Wherever possible, reviews should be done blind, so the reviewer does not know whom they are reviewing. A well-known example of the effectiveness of this technique is in orchestra auditions, where the proportion of women hired shot up when auditions were performed anonymously behind a curtain.

The literature abounds with other best practices for academia (see the United Kingdom's Athena SWAN Charter or the US National Science Foundation's ADVANCE programme). What is missing is not ways to do better — but the recognition that we must change.

Different ideas lead to scientific advances. Rome projected influence over a great empire, but did not foster a distinguished scientific enterprise: the greatest discoveries tended to come at the intersections of trade routes. Sameness leads to stagnation. We simply have to try. Harder. ■ [SEE NEWS REVIEW P.451](#)

Meg Urry is professor of physics and astronomy, and director of the Yale Center for Astronomy & Astrophysics, Yale University, New Haven, Connecticut, USA. e-mail: meg.urry@yale.edu

1. Mason, M. A. & Goulden, M. *Academe* **88**, 21–27 (2002).
2. West, J. D., Jacquet, J., King, M. M., Correll, S. J. & Bergstrom, C. T. *PLoS ONE* **8**, e66212 (2013).
3. Larivière, V., Ni, C., Gingras, Y., Cronin, B. & Sugimoto, C. R. *Nature* **504**, 211–213 (2013).
4. Pohlhaus, J. R. et al. *Acad. Med.* **86**, 759–767 (2011).
5. Moss-Racusina, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J. & Handelsman, J. *Proc. Natl Acad. Sci. USA* **109**, 16474–16479 (2012).
6. Trix, F. & Penska, C. *Discourse Soc.* **14**, 191–220 (2003).
7. Reid, N. I. *Publ. Astron. Soc. Pacific* **126**, 923–934 (2014).
8. Lincoln, A. E., Pincus, S., Bandows Koster, J. & Leboy, P. S. *Soc. Stud. Sci.* **42**, 307–332 (2012).
9. Shipman, C. & Kay, K. *The Confidence Code* (HarperBusiness, 2014).



George Schaller looks for Marco Polo sheep (*Ovis ammon polii*) in Afghanistan in 2004.

CONSERVATION BIOLOGY

Wild at heart

Henry Nicholls talks to pioneering field biologist George Schaller — still studying iconic species at 82.

I sit in the forest. I hear branches crackling,” says George Schaller, recalling a close encounter with a wild giant panda more than 30 years ago. The large female sits down just 5 metres from him. “Her head sinks to her chest and she falls asleep,” he says.

With a career spanning more than six decades, pioneering field biologist Schaller is no stranger to such moments. He made the first studies of an extraordinary range of charismatic mammals, including the Bengal tiger (*Panthera tigris tigris*), the East African lion (*Panthera leo nubica*), the snow leopard (*Panthera uncia*) and the Tibetan antelope, or chiru (*Pantholops hodgsonii*), as well as the giant panda (*Ailuropoda melanoleuca*). He has tracked some seriously elusive mammals, confirming the existence of the antelope-like saola (*Pseudoryx nghetinhensis*) in Laos and locating a new population of Tibetan red deer (*Cervus canadensis wallichi*) not far from Lhasa. He has distilled the essence of hundreds of such sojourns in the wildest regions on Earth into 15 books, 7 of them intended for an academic audience. These include

The Giant Pandas of Wolong (University of Chicago Press), coauthored with Hu Jinchu, Pan Wenshi and Zhu Jing 30 years ago.

Schaller's first overseas expedition left the United States for what is now the Democratic Republic of the Congo in 1959, to study the mountain gorilla (*Gorilla beringei beringei*) in the Virunga Mountains. Sponsored by the New York Zoological Society, this was the first serious scientific study of the species, and it paved the way for the work of primatologist Dian Fossey. On one occasion, he climbed a tree to get a better view of a gorilla family and was joined on a branch by one of the females. “We were both nervous, but something like that never leaves you,” he says. Now 82, he returned in September from a five-week expedition to Brazil in his capacity as vice-president of wild-cat conservation charity Panthera in New York. The group's ambitious Jaguar Corridor Initiative seeks to protect the species across its entire 6-million-square-kilometre range, which spans 18 countries from Mexico to Argentina.

The highlight of the trip, says Schaller, was a rare sighting in Brazil's Amazonia National

Park in the southwest of the Amazon basin. “We were on a small boat in one of the rivers and spotted a beautiful black jaguar, all glossy with muted gold eyes,” he says. “It lay on the bank of the river. We watched it for half an hour and then left just to give it peace.”

As Schaller has got to know individual animals and species arguably better than anyone else alive, he has advocated tirelessly for their protection. Few policymakers or members of the public “read scientific papers or could care less about them,” he points out. So as well as writing papers and academic studies, and making recommendations to government departments, he has devoted considerable energy to making his findings accessible through popular-science books, eight so far. The public responds to the plight of large, often charismatic animals. “That automatically provides protection to all the other species and the habitats in that area,” he says.

When he began the gorilla study, Schaller planned to write two books, one technical and one popular (they became, respectively, *The Mountain Gorilla* (1963) and *The Year of the Gorilla* (1964), both published by University of Chicago Press). He carried two notebooks into the rainforest: one for field notes, the other for personal reflections. “Memory is lousy,” he says. He has stuck to this technique, despite pens frozen in sub-zero temperatures and lampless tents filled with wood smoke. His devotion to the daily ritual of transcribing his thoughts and emotions shows in the dramatic first-hand detail that defines much of his writing.

“If you look at nature shows on television, most of them are dismal,” says Schaller. “Beautiful animals, but no message.” His popular books offer a direct challenge to such simplistic visions. They do focus on stunning landscapes and the fascination of individual animals and iconic species, but Schaller offers more. Successful conservation, to succeed, must operate in a complex

“Most nature shows on television are dismal. Beautiful animals, but no message.”

cultural ecosystem, as important to Schaller as natural ecosystems. And decades of field-work have given Schaller a talent for observing humans — creatures who are, he says, “much better at hiding their real actions and thoughts than animals are”.

There is a risk in telling it how it really is: “In some countries, if you say too much you can't go back.” His critique of individual and institutional failings in *The Last Panda* (University of Chicago Press, 1993) “got some interesting responses”.

But he clearly didn't go too far. Schaller has spent more time in China than in any other country. After *The Giant Pandas of Wolong* was published in 1985, he left the study of

the species to his Chinese colleagues. Yet he is drawn back year after year to the Changtang, the great northern plain on the Tibetan Plateau, to study species such as the chiru, the wild ass called the kiang (*Equus kiang*) and the wild yak (*Bos mutus*), as well as snow leopards. *Tibet Wild* (Island, 2012) chronicles the challenges and joys of conducting research on Earth's highest plateau.

"One reason I like working in China is that the people are very pragmatic," he says. He is only just back from participating in a snow-leopard survey on the plateau, where winter temperatures frequently fall below -30°C . In the new year, he is off to Iran to check on the Asiatic cheetah (*Acinonyx jubatus venaticus*).

Given that Schaller has witnessed the destruction of habitats, the fragmentation of populations and the trade in endangered species, is he disillusioned? Although he acknowledges that apathy, greed and corruption threaten nature, he recognizes major achievements. The population of mountain gorillas has recovered to roughly where it was around 50 years ago; China has created more than 60 national parks across the giant panda's range; the illegal poaching of chiru

G. SCHALLER



Schaller, a herdsman and a snow leopard.

for their fur has been brought under some control in China. The Changtang Nature Reserve, established in 1993 as a direct result of Schaller's work, is larger than Italy.

Schaller's legacy also has a strong human dimension. "The thing I treasure most is leaving behind young biologists who worked with me and who will carry on to train the next generation," he says. "I get uplifted all the time. I see the progress." ■

Henry Nicholls is author of *The Way of the Panda* and the *Animal Magic* blog at *The Guardian*.

e-mail: henry@henrynicholls.com

Books in brief



Searching for the Oldest Stars: Ancient Relics from the Early Universe

Anna Frebel (translated by Ann M. Hentschel) PRINCETON UNIVERSITY PRESS (2015)

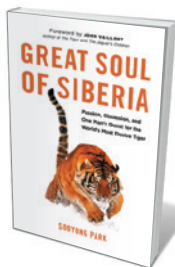
As a "stellar archaeologist", Anna Frebel tracks metal-poor stars — the "ancient messengers" that kick-started the cosmos's chemical evolution. Her discoveries include a Milky Way star 13.2 billion years old and superannuated stars in dwarf galaxies that orbit our own. In this account of her work, she neatly balances the technical and the personal — not least in chapters on the mesmerizing slog of nightly observations, many using Chile's 6.5-metre Magellan telescopes.



Patternalia

Jude Stewart BLOOMSBURY (2015)

We are often only half-aware of graphic patterns such as paisley or polka dots, or the patterns that pulsate in nature, from fractals to flocking birds. Jude Stewart here brings "patternalia" to the fore and crisply decodes the mathematical, scientific and cultural connotations behind it. Dip in for some pointed erudition on the tension between comforting algebraic numbers and their 'transcendental', patternless cousins; varieties of military camouflage from chocolate chip to tiger stripe; and the revolution wrought by the programmable, futuristic Jacquard loom, demonstrated in 1801.



Great Soul of Siberia: Passion, Obsession, and One Man's Quest for the World's Most Elusive Tiger

Sooyong Park GREYSTONE (2015)

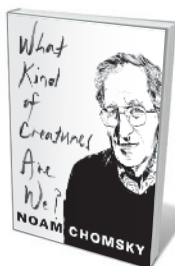
Just 350 Siberian tigers from a once thousands-strong population pad through Russia's northeastern birch forests: massive, elusive, "burning bright". For this astonishing ethological study, South Korean film-maker Sooyong Park spent two decades alternately tracking the beasts and holed up in underground bunkers, seeking glimpses of them in subzero weather. His paean to one of the world's biggest cats has a piercing immediacy distilled from thousands of heart-stopping sightings and encounters. A landmark achievement.



First Bite: How We Learn to Eat

Bee Wilson BASIC (2015)

With televised cake-baking compulsive viewing and Western obesity levels at an all-time high, humanity's relationship with food is a strange melange. For her lucid survey, journalist Bee Wilson uses how we eat as children as a springboard for discussions of the wilder shores of adult consumption. Along the way, she dishes up an impressive range of research in neuroscience and nutrition on topics from the evolution of the Japanese diet to babies' self-directed preferences for, say, turnips, as demonstrated in the fascinating, flawed work of twentieth-century US paediatrician Clara Davis.



What Kind of Creatures Are We?

Noam Chomsky COLUMBIA UNIVERSITY PRESS (2015)

At 87, linguist Noam Chomsky is still nimbly tackling big questions about human nature — here, in less than 200 pages. Hanging his analysis off palaeontologist Ian Tattersall's theory that the human sensibility was born 50,000–100,000 years ago, he remakes his case for biology-based linguistics, discusses the "new mysterianism" that is delimiting humanity's capacity for comprehension, and extols libertarian socialism. However, although thoughtful individually, these arguments betray their origins as lectures and fail to gel. [Barbara Kiser](#)

MATHEMATICS

Geometries of beauty

Lynn Gamwell traces the millennia of symbiosis between mathematics and art.



Sandro Botticelli depicts Saint Augustine with mathematical trappings such as an armillary sphere.

Throughout history, mathematics has developed as part of humanity's search for patterns. When I explored mathematics in cultures East and West for my book *Mathematics and Art: A Cultural History* (Princeton University Press, 2016), I discovered that many artists have expressed their cultural world views through works that embody these patterns. From the classical period more than two millennia ago,

through the Chinese dynasties, the Western Renaissance and the mathematics and physics of the twentieth and twenty-first centuries, art and architecture have incorporated the mathematics of their day in deep ways.

How a culture conceives of ultimate reality — whether it is composed of atoms or a 'world soul' — relates directly to how people think about mathematics. Plato looked to pure form: abstractions such as numbers or

spheres that resided outside mundane time and space. He described a divine being, the demiurge (from the Greek for 'craftsman'), who created the natural world by imposing these archetypal forms onto formless matter. The thinker Augustine of Hippo in North Africa later bridged classical knowledge and Christian theology in *De Doctrina Christiana* (around AD 400). Augustine — who had studied the seven liberal arts, including geometry and astronomy, before his conversion to Christianity — noted that if the Platonists "have said things which are indeed true and well accommodated to our faith, they should not be feared".

More than a millennium later, that mingling of Hellenic and Christian thought was expressed by Italian Renaissance artist Sandro Botticelli. In a 1480 fresco in the Ognissanti church in Florence, Botticelli portrayed Augustine as a scholar-saint wearing clerical robes with an open treatise on geometry and a weight-driven clock nearby. He looks heavenwards, seeking the order that the Christian God (like Plato's demiurge) imposed on creation by dividing light from darkness. But Botticelli's rendering of Augustine also reveals the influence of classical Greco-Egyptian mathematician Ptolemy, who reasoned that in the ideal perfection of the heavens, bodies such as planets are spherical and move in circular paths at a constant speed. An Earth-centred Ptolemaic armillary sphere appears in the upper left-hand corner of the fresco.

Botticelli's mathematical awareness extended to an understanding of linear perspective: he portrays Augustine as a solid body inhabiting an architectural space. Decades earlier, his fellow Florentine Filippo Brunelleschi had invented a way to visualize a geometric projection from a given viewpoint, based on findings in optics by medieval Islamic scholar Ibn al-Haytham — the first to explain vision as the eye's passive response to light (J. Al-Khalili *Nature* **518**, 164–165; 2015). Brunelleschi's experimentation allowed Florentine artists to paint figures not floating in a golden mist, but right here, right now, in a believably depicted natural world.

Ancient Greek mathematicians abstracted and generalized from the particular to the whole; ancient Chinese mathematicians did not generalize, but focused on particular examples as paradigms. The Chinese were adept at creating numerical patterns such as the Luoshu, in which even-odd, black-white

ELECTA/MONDADORI PORTFOLIO/THE ART ARCHIVE



Czech Modernism Mirrored and Reflected Infinitely (2005) by Josiah McElheny reflects an enclosed world to infinity.

or yin–yang pairs of numbers symbolizing the elements (metal, fire, water, wood, earth) are arranged so that the rows, columns and diagonals add up to the same number. This ‘magic square’ reflects the Taoist view that the natural world is a balance of parts, which came into being by self-assembly following the ultimately unknowable Tao (way) of nature. Japanese artist Tatsuo Miyajima echoes this outlook in his 1998 artwork *Keep Changing, Connect with Everything, Continue Forever*, a sparkling, blinking grid of red light-emitting diodes. The overall pattern is too complex for the human eye and mind to discern, but, like the natural world, it is a unity of fluctuating parts.

While Chinese mathematicians studied dynamic patterns in harmonic balance, Western mathematics rested on a rock-solid foundation put in place by Euclid, a follower of Plato, in his treatise *Elements* (around 300 BC). Euclid organized certain truths (for example, that all right angles are equal), into an axiomatic system that undergirded Western mathematics until the nineteenth century, when other geometries were discovered. In his *Foundations of Geometry* (1899), the German mathematician David Hilbert cut the ties that bound points and planes to the world, rewriting geometry as an internally consistent arrangement of meaning-free signs. Inspired by Hilbert’s formalist approach, the Russian artist Aleksandr Rodchenko reduced painting to its essence — a monochrome rectangle — in works such as

Red (1921). This aesthetic remained at the core of nonobjective art throughout the twentieth century.

The reductionist impulse survives today in artists such as Josiah McElheny, who, echoing Hilbert, created a self-contained arrangement of abstract signs, *Czech Modernism Mirrored and Reflected Infinitely* (2005). McElheny transformed eight glass bottles into mirrors with an interior coating of silver. Set within a mirrored box, they reflect all light to infinity. Like Hilbert’s mute marks, they are uncontaminated in their self-enclosed world: the front of the box is a one-way mirror, reflective surface facing inwards.

In *Mathematics and Art*, I also trace the tension between accounts of deterministic laws of cause and effect in nature, and rebellions against the ‘dehumanizing’ nature of such laws and their associated mathematics. In antiquity, the Greek rationalist Democritus described a mechanical, predictable Universe made of inert atoms. But in Plato’s cosmos, humanity’s attainment of certain knowledge was unpredictable: only after lengthy contemplation “does truth flash upon the soul, like a flame kindled by a leaping spark”.

In the twentieth century, this tension was expressed in rival philosophical interpretations of the subatomic realm of quantum physics. The leading contender was the Copenhagen interpretation, put forth in 1927 by Niels Bohr and Werner Heisenberg, who declared that nature is fundamentally indeterministic and reality is in the mind of the observer. The minority view held out for universal knowledge and a physical world

“Art and architecture have incorporated the mathematics of their day in deep ways.”

independent of human observation. Its spokesman, an exasperated Albert Einstein, exclaimed: “The moon is there even when I’m not looking at it.” After the Second World War, Bohr and Heisenberg expressed the Copenhagen interpretation in popular science writings (such as Heisenberg’s *Physics and Philosophy*, 1959), in which they announced that the classical ideals of rationality and objectivity were naive. Such declarations contributed to postmodernism, a stance widely adopted by artists.

Ultimately, mathematicians and artists often hybridize Western and Eastern traditions, which began to merge after Charles Darwin published *On the Origin of Species* in 1859. As Western theology made way for science, more and more features of the human body and mind yielded to the explanatory power of biology, physiology and psychology. This cataclysmic shift prompted many Westerners to integrate into their work the Taoist view of nature as a balance of parts coming together by self-assembly.

The scientific world view is a hybrid of Western and Eastern traditions, expressing the philosophical conviction that the natural world has a wholeness with which humans are one. Today, there are new manifestations of the quest for unity. Scientists and mathematicians continue to search for patterns in the physical Universe that in turn inspire artists who strive to express our global cultural view of reality. ■

Lynn Gamwell is a lecturer on the history of art, mathematics and science at the School of Visual Arts in New York. Her previous books include *Exploring the Invisible*.
e-mail: lgamwell@sva.edu

Correspondence

Spider taxonomists catch data on web

ILLUSTRATION BY PHIL DISLEY

A successful systematics initiative in arachnology could provide an invaluable model for rapid delivery of taxonomic data for other animal groups. Until now, the inaccessibility of the classical and obscure taxonomic literature has been a major hindrance to the field's progress.

The World Spider Catalog (www.wsc.nmbe.ch), launched last year, now contains complete taxonomic data for almost 46,000 validated spider species and an embedded collection of 13,000 references. Spiders are the most species-rich terrestrial invertebrate group after insects.

More than 97% of the world's spider literature was collected within just 600 days of communicating our goal to the research community. The database logs a daily average of 600 hits and 400 downloads.

Wolfgang Nentwig *University of Bern, Switzerland.*

Daniel Gloor *Natural History Museum Bern, Switzerland.*

Christian Kropf *University of Bern; and Natural History Museum Bern, Switzerland.*
wolfgang.nentwig@iee.unibe.ch

Bury the idea that soils are a local issue

As the International Year of Soils ends, we agree that the importance of integrating soils into policies to tackle global challenges cannot be underestimated (see L. Montanarella *Nature* **528**, 32–33; 2015). Soils are not a local issue — they 'move' at time and space scales that are relevant to global policy.

For example, Saharan soil dust has boosted Atlantic plankton (E. Marañón *et al. Limnol. Oceanogr.* **55**, 2339–2352; 2010) and tree growth in Amazonian forests (R. Swap *et al. Tellus B* **44**, 133–149; 1992). There are environmental consequences beyond national borders when pollutants and nutrients that are attached to soil particles



enter waterways, or when soil nitrates leach into aquifers. The influence of soil management on climate is also global because of its carbon-storage capacity and interactions with greenhouse gases. Changes in soil-surface reflection (albedo) affect energy balance, climate and weather.

Frank G. A. Verheijen, Ana C. Bastos *University of Aveiro, Portugal.*

Simon Jeffery Harper Adams *University, Newport, UK.*
frankverheijen@gmail.com

Labs should cut plastic waste too

Many governments now impose charges for single-use plastic bags and bottles. As responsible researchers, we should cut back on disposable plastics (see also G. Bistulfi *Nature* **502**, 170; 2013).

We estimate that the 280 bench scientists in our bioscience department generated roughly 267 tonnes of plastic in 2014 (data from University of Exeter Sustainability and Waste and Resource Management offices). That is equivalent to about 5.7 million empty 2-litre plastic bottles. Some 20,500 institutions worldwide are involved in biological, medical or agricultural research (where plastic disposal is likely to be heaviest), so that could equate

to around 5.5 million tonnes of lab plastic waste in 2014 — roughly the combined tonnage of 67 cruise liners, and equal to 83% of the plastic recycled worldwide in 2012.

We justify our use of disposables on the grounds of costs and time saved. Grant agencies therefore need to introduce incentives to reduce plastic waste, for example by funding lab washing-up and recycling facilities, and possibly to make greener lab practices a requirement in the grant-application process.

Mauricio A. Urbina *University of Exeter, UK; and University of Concepción, Chile.*

Andrew J. R. Watts, Erin E. Reardon *University of Exeter, UK.*
mauriciourbina@udec.cl

Nuclear industry no model for biosafety

I applaud Tim Trevan's call to reform lab biosafety, but disagree with his argument for using the nuclear industry as a model (*Nature* **527**, 155–158; 2015).

Nuclear facilities are strictly regulated and ensure that potential hazards arising from process changes are engineered out (see go.nature.com/qyzoth). Yet scientists are not process-driven: being autonomous and creative, they need freedom

to change and require a dynamic safety culture that can accommodate new challenges. These include the replacement of humans by technology, reduced supervision and declining safety competencies — none of which applies to the nuclear industry.

Chasing a 'zero harm' mantra can actually promote a poor safety culture because it is an outcome rather than a goal (for examples of alternative approaches, see go.nature.com/xgupio and go.nature.com/gcjqlf). As Trevan points out, an effective safety culture is measured through engagement, understanding and care for everyone's well-being.

Chris Lea *UCB Celltech, Slough, UK.*
chris.lea@ucb.com

Bond villain fails neuroanatomy

The thrills and action in *Spectre*, the latest James Bond film, were somewhat marred for this viewer by a fundamental neuroanatomical blunder.

The scene is a Moroccan desert. Bond's nemesis is torturing our hero using a head clamp fused with a robotic drill. Intending to erase Bond's memory of faces, the villain says he is directing his drill to the (lateral) "fusiform gyrus" — correctly identifying a core brain area for facial recognition (J. Parvizi *et al. J. Neurosci.* **32**, 14915–14920; 2012).

But the film-makers got it wrong. Whereas the drill should have been aimed just in front of 007's ear, it was directed below the mastoid process under and behind his left ear. There it would have met the lateral part of the first or second cervical vertebra, perhaps hitting the ipsilateral vertebral artery and triggering a stroke or massive haemorrhage. Unless fatal, it certainly would not have deleted the bank of faces in Bond's memory.

Michael D. Cusimano *St Michael's Hospital, University of Toronto, Canada.*
mountain@smh.ca

Maurice Strong

(1929–2015)

Oil man who was first director of the United Nations Environment Programme.

That anthropogenic climate change is now of mainstream concern has, paradoxically, a lot to do with an oil man. Maurice Frederick Strong, fossil-fuel magnate, was the founding executive director of the United Nations Environment Programme (UNEP). He died on 27 November.

Strong was among the last of a generation of post-war scholar-administrators in Canada that included former prime minister Lester B. Pearson. Each lived through the Great Depression and the Second World War, determined that history must never be repeated. Pearson positioned Canada as the world's anti-poverty champion. Strong was one of Pearson's army of nation-builders: he helped to create the Canadian International Development Agency in 1968 and the national oil company Petro-Canada in 1976.

It is thanks to UNEP that nearly every government today has a dedicated department that looks after the environment. The body's creation in 1972 can be attributed directly to Strong's unusual blend of skills. He was adept at making complex science accessible to non-specialists, and able to build unlikely coalitions to support his cause. In 2009, he summarized his approach as, "never to confront, but to co-opt, never to bully but to equivocate, and never to yield".

Strong was born in April 1929 in Oak Lake, rural Canada, to a family that had fallen on desperate times. As he wrote movingly in his autobiography, *Where on Earth Are We Going?* (Knopf, 2000), the Great Depression "stripped my father of his livelihood and his sense of self worth. It ruined my mother's health and in the end it killed her". In winter their clothes would freeze stiff, and at times there was little to eat beyond weeds and dandelions. The need and hunger he witnessed haunted him for years.

Leaving school in 1943, Strong won a cash prize to help with the costs of university but used the money to pay off his parents' creditors. He did not join the ranks of young men headed for the front line. Waiting to board a freight train near his home, he spotted a discarded copy of the local newspaper. He read that Winston Churchill and Franklin D. Roosevelt had decided that once the war was over they would work to unite nations. Strong decided that he wanted to be a part of that project.

For the next two decades he forged what many saw as a contradictory career: that of an oil tycoon. He built companies. Bought



companies. Sold companies. He acquired unparalleled knowledge and experience of the energy business; and he became rich. Oil wealth for Strong had a second purpose: it was his passport to Canada's elites. His prominence and ability to make money caught the attention of ministers, and that enabled Strong to realize his public-service ambitions, at home and on the world stage.

In 1969, Strong was running Canada's aid programme when Sweden sought his advice on how to rescue a global environmental meeting. The conference was due to take place in Stockholm in 1972. Few wanted to come, and many of the nations that had signed up seemed to wish the conference to fail.

Developed countries were yet to be convinced that the environmental threat was real. Solly Zuckerman, a former chief scientific adviser to the British government, branded Strong an "extremist", claiming that environmental degradation was reversible.

Developing nations had different concerns. Some with ambitions to industrialize saw the conference as a conspiracy to keep them poor. "The Third World is not merely worried about the quality of life, it is worried about life itself," said Pakistan's former chief economist Mahbub ul Haq. And the Soviet bloc was threatening a boycott because the United States wanted communist East Germany excluded from the meeting.

Strong set to work. As the conference secretary-general, he appointed a Soviet

scientist to his staff, which gave him a direct line to negotiate with Moscow. And he asked the developing countries to set the conference agenda. This would say explicitly that they could protect their environments without compromising their ability to industrialize, and that rich countries should help to finance poor countries to achieve that goal. To deal with objections from British scientists, Strong sought help from the team at the Massachusetts Institute of Technology in Cambridge that had just published the book *The Limits to Growth* (Universe, 1972).

He put Barbara Ward, former foreign editor of *The Economist* turned environmental advocate, on the conference staff to neutralize the influence of sceptical diplomats from rich nations. Somehow, Strong persuaded Indira Gandhi, then-prime minister of India, to open the conference.

The 1972 United Nations Conference on the Human Environment ended in practical action. It led to a new UN body to monitor the global environment, to be based in Nairobi, Kenya. Strong remained UNEP's executive director until 1975. Two decades later, the UN called on him again to steer the Earth Summit in Rio de Janeiro and this resulted in three further agreements: the Framework Convention on Climate Change, the Convention on Biological Diversity and, later, the Convention to Combat Desertification.

Yet Strong's diplomatic ability was not universally appreciated. Many in the energy industry saw him as a closet 'green'; to environmental groups he represented Big Oil. The right, meanwhile, attacked him as the embodiment of Big Government. It is true that UNEP and the environment conventions have made little progress in slowing climate change or reducing the rate of biodiversity loss.

Such failures cannot be attributed to Strong alone. They point to a flaw in the global environmental architecture that he helped to draw up. The world's green agreements need leaders with an unusually broad mix of qualities. Maurice Strong was one of the last. His passing is the end of an era. ■

Ehsan Masood is editor of *Research Fortnight*. His next book is *The Great Invention: The Story of GDP and the Making and Unmaking of the Modern World*.

e-mail: ehsan.masood@researchresearch.com
Twitter @ehsanmasood

KEYSTONE PICTURES USA/ALAMY STOCK PHOTO

The observation of square ice in graphene questioned

ARISING FROM G. Algara-Siller *et al.* *Nature* **519**, 443–445 (2015); doi:10.1038/nature14295

Algara-Siller *et al.*¹ reported the observation of a new phase of water—‘square ice’—sandwiched between two graphene layers at room temperature. Their key evidence consists of transmission electron microscope (TEM) images of a square lattice from small encapsulated crystals, the detection of oxygen from relatively large regions containing such crystals and molecular dynamics (MD) simulations indicating ‘square ice’ formation inside hydrophobic nanochannels. Here we propose that the reported experimental data can be better explained by salt (for example, NaCl) contaminants precipitating as nanocrystals in the dried-out graphene liquid cells² and common oxide contaminants in graphene. Consequently, we question the observation of room-temperature ‘square ice’. There is a Reply to this Brief Communication

Arising by Algara-Siller, G. *et al.* *Nature* **528**, <http://dx.doi.org/10.1038/nature16149> (2015) relating to the electron energy-loss spectra (EELS) and a Reply by Wang, F. C. *et al.* *Nature* **528**, <http://dx.doi.org/10.1038/nature16146> (2015) relating to the MD simulations.

The TEM images and the dynamics of the reported ‘square ice’ under electron irradiation bear a considerable resemblance to those we have observed of NaCl nanoplatelets in graphene and dried-out graphene liquid cells. Such NaCl nanoplatelets usually orient along the [100] direction, displaying a square lattice with a spacing of approximately 2.8 Å (Fig. 1a, b, d); the corresponding fast Fourier transform (FFT) matches the reported¹ diffraction data. Edge termination, dislocations and grain-boundary structures within the lattice, and the dynamics of

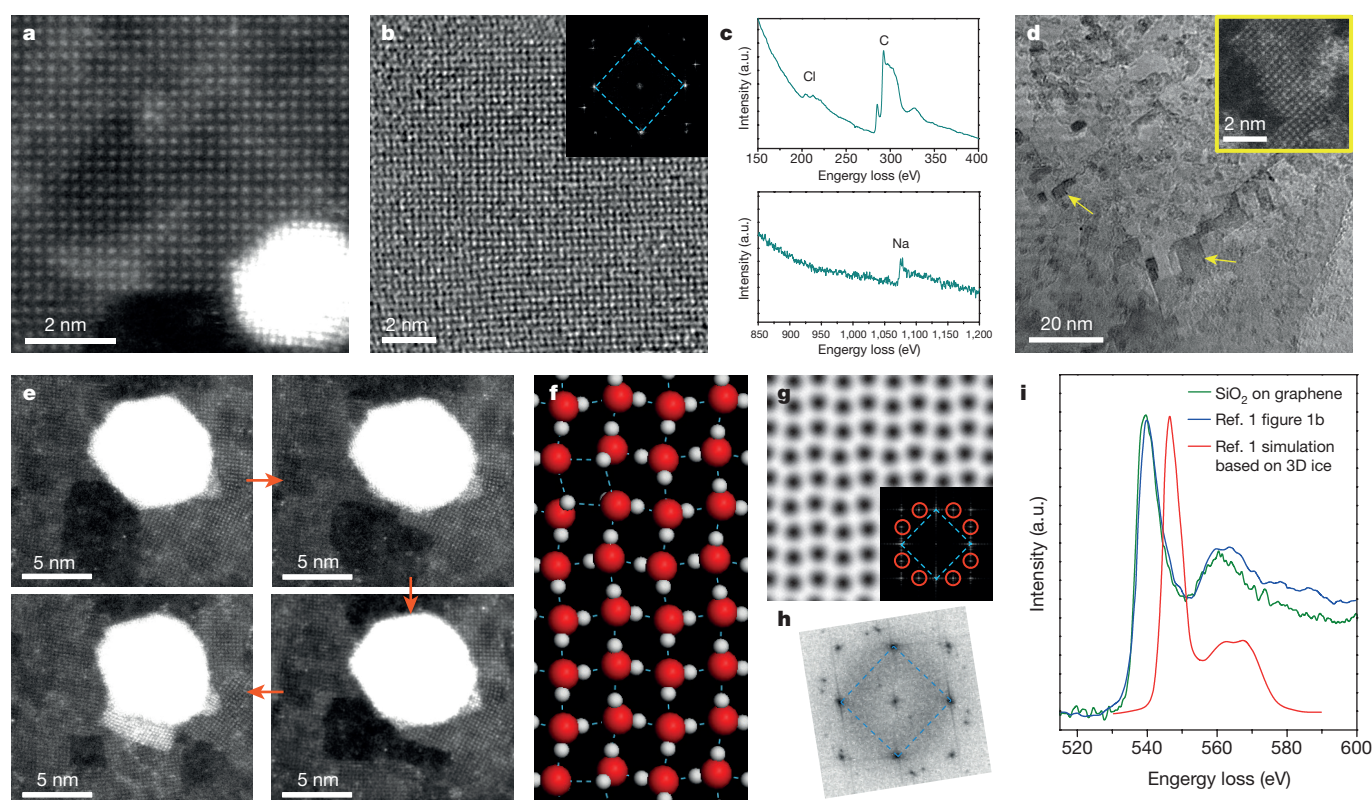


Figure 1 | Structure and analysis of NaCl nanoplatelets and reported ‘square ice’. **a**, Scanning TEM annular dark field (STEM-ADF) image of a NaCl nanoplatelet in graphene. **b**, STEM bright field image (main panel) and the FFT (inset) of a NaCl nanoplatelet. The blue dashed lines highlight the four equivalent {200} planes with 90° interplanar angle. **c**, EELS acquired from a 4 nm × 4 nm region of the NaCl nanoplatelet. The peaks corresponding to chlorine (Cl), carbon (C, from graphene) and sodium (Na) are labelled. a.u., arbitrary units. **d**, TEM bright field image of a dried-out graphene liquid cell containing NaCl residuals (indicated by the yellow arrows) that exhibit chequered patterns similar to those reported in ref. 1. Inset, ADF image taken from a small region of a typical chequered pattern. **e**, Sequential (indicated by the red arrows) STEM-ADF images showing the reconstruction of NaCl crystals under electron beam irradiation, as

seen by the constant change of the outline of the thin NaCl crystal. **f**, Ice structure in the graphene layers calculated using DFT. Graphene layers are not shown. Red circles denote oxygen; white circles denote hydrogen; blue dashed lines represent hydrogen bonds. **g**, Simulated image (main panel) and the corresponding FFT (inset) of monolayer ‘square ice’ based on the structure calculated using DFT. The red circles highlight the additional diffraction spots arising from the rhombic, zig-zag structure; the blue dashed lines highlight the square structure. **h**, FFT adapted from figure 1a in ref. 1; the blue dashed lines highlight the square structure. **i**, Comparison of the experimental oxygen K-edge EELS from ‘2D ice’ (ref. 1) and the corresponding simulated EELS (from 3D ice, adapted from ref. 9 and reported in ref. 1) with our experimental oxygen K-edge EELS from SiO_x particles on graphene.

NaCl nanoplatelets under electron irradiation (Fig. 1e), also resemble what is reported in ref. 1. EELS collected from $4\text{ nm} \times 4\text{ nm}$ regions inside the nanocrystals have clear chlorine and sodium peaks (Fig. 1c) and indicate the presence of a trace amount of oxygen due to contamination, confirming that these crystals are NaCl.

Although the MD simulations of monolayer water molecules in graphene nanocapillaries reveal a structure similar to that seen in the TEM images of ref. 1, the simulated graphene-confined 2D ice crystals do not have a perfectly square structure, but instead exhibit a zig-zag arrangement of water molecules (figure 2d in ref. 1). In our density functional theory (DFT) calculations (Fig. 1f) and MD simulations, we find that densely packed water molecules inserted between two graphene sheets as a perfect square lattice undergo relaxation to yield a rhombic, zig-zag structure. This rhombic structure, found previously in simulations of 2D water-molecule structures^{3,4}, is sometimes referred to as ‘square ice’ to indicate the fourfold coordination of water molecules as opposed to the three-fold coordination in conventional ice; to our knowledge, ‘square ice’ has not previously been used to indicate square symmetry^{3,4}.

The zig-zag structure reduces the symmetry and generates additional spots in the electron diffraction pattern and FFT (Fig. 1g). The FFT of the reported¹ TEM image (taken over 1 s) lacks these extra spots (Fig. 1h), indicating that the image was generated from a crystal with higher apparent symmetry than that of the ice structure presented in the reported¹ MD snapshot (taken over 1 fs) or seen in our DFT and MD calculations (Fig. 1f). It might be argued that lattice vibrations average out the positions of the oxygen atoms to yield a square lattice; however, the averaged oxygen positions in our MD snapshots taken over 200 ps still have a rhombic structure. In the absence of a demonstration that averaging over macroscopic timescales produces a square lattice, the reported¹ MD simulations do not seem to support the observation of ‘square ice’.

We conclude that crystals with a highly symmetric rock-salt structure⁵ and atomic column spacing of 2.82 Å , such as NaCl, better account for the TEM images and the corresponding diffraction data presented in ref. 1 than does ‘square ice’; the NaCl structure also explains the observed stacking at larger crystal thickness that the reported¹ MD simulation was unable to reproduce.

An experimental oxygen K-edge EELS taken from an area with a diameter of about 100 nm containing the much smaller ‘square ice’ crystals (about 10 nm in diameter) was also compared¹ to a calculated oxygen K-edge EELS of 3D ice, revealing the spectra to be “qualitatively similar”, but with a peak shift of approximately 6 eV. The differences were attributed¹ to those between 3D and ‘2D square ice’; however, Fig. 1i demonstrates that the reported¹ experimental oxygen K-edge EELS is in nearly perfect agreement with the oxygen K-edge spectrum of SiO₂, a common contaminant in graphene samples. Given that Algara-Siller *et al.*¹ reported the presence of silicon EELS signals throughout their sample, the reported oxygen signal might well have a large contribution from contaminants and thus should not be considered evidence of ice. Furthermore, the EELS data provided in ref. 1 do not rule out the presence of other elements such as sodium and chlorine, which have weak edges or edges outside typical spectrum ranges.

In conclusion, we believe that the experimental data presented by Algara-Siller *et al.*¹ neither provide definitive evidence for the existence of ‘square ice’, nor agree with their reported theory. We suggest that accidental contamination with NaCl (or another salt with similar structure) and subsequent salt nanocrystal formation better explains the reported experimental data than does ‘square ice’. Salts are hygroscopic and so would be associated with any water left in the graphene liquid cells and would precipitate upon water evaporation under electron beam irradiation. Further experimental and theoretical studies are required to assess the existence of ‘square ice’ in graphene nanochannels.

Methods

Graphene liquid cells containing 0.06 M NaCl solution were prepared using the method described in ref. 2. The liquid cells were dried out inside the TEM by exposing them to high-electron-dose-rate illumination, which generates bubbles and causes the liquid cells to burst. The dried-out cells prepared with dilute NaCl solution have a considerable amount of visible residue, whereas graphene liquid cells prepared with de-ionized water have no visible residue after drying.

High-resolution TEM image simulation (Fig. 1g) was performed on our DFT-calculated structure using the MacTempasX software (<http://www.totalresolution.com/MacTempasX.htm>) and on the structural model reported in the MD snapshot (figure 2d in ref. 1) using codes provided in ref. 6; very similar results were obtained. Parameters for image simulation are: accelerating voltage, 80 kV; spherical aberration C_s , 30 μm ; defocus, -13.4 nm .

Both *ab initio* MD and classic MD simulations were performed to test whether the zig-zag structure of the water molecules could be smoothed to yield a square pattern over a certain length of time. For the *ab initio* MD simulation with a PBE (Perdew–Burke–Ernzerhof) functional⁷, 36 water molecules were constrained between two graphene layers at a higher density than that of water at 4°C and 1 atm (and similar to that in ref. 3). The average positions of oxygen atoms over 2 ps (2,000 timesteps) remain in a rhombic pattern. The classic MD simulation was run using the COMPASS force field⁸, and 144 water molecules were constrained between two graphene layers. One oxygen atom was fixed to avoid a translational motion. The overall trajectories of oxygen atoms over 200 ps (200,000 timesteps) still have a noticeable zig-zag structure.

Wu Zhou¹, Kuibo Yin^{1,2}, Canhui Wang³, Yuyang Zhang^{1,4}, Tao Xu², Albina Borisevich¹, Litao Sun², Juan Carlos Idrobo⁵, Matthew F. Chisholm¹, Sokrates T. Pantelides^{1,4}, Robert F. Klie³ & Andrew R. Lupini¹

¹Materials Science & Technology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA.

email: wu.zhou.stem@gmail.com

²SEU-FEI Nano-Pico Center, Key Lab of MEMS of Ministry of Education, Southeast University, Nanjing 210096, China.

³Department of Physics, University of Illinois at Chicago, Chicago, Illinois 60607, USA.

⁴Department of Physics and Astronomy, Vanderbilt University, Nashville, Tennessee 37235, USA.

⁵Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA.

Received 5 May; accepted 22 September 2015.

1. Algara-Siller, G. *et al.* Square ice in graphene nanocapillaries. *Nature* **519**, 443–445 (2015).
2. Wang, C., Qiao, Q., Shokuhfar, T. & Klie, R. F. High-resolution electron microscopy and spectroscopy of ferritin in biocompatible graphene liquid cells and graphene sandwiches. *Adv. Mater.* **26**, 3410–3414 (2014).
3. Zhao, W., Bai, J., Yuan, L., Yang, J. & Zeng, X. Ferroelectric hexagonal and rhombic monolayer ice phases. *Chem. Sci.* **5**, 1757–1764 (2014).
4. Takaiwa, D., Hatano, I., Koga, K. & Tanaka, H. Phase diagram of water in carbon nanotubes. *Proc. Natl Acad. Sci. USA* **105**, 39–43 (2008).
5. Bragg, W. H. & Bragg, W. L. The reflection of X-rays by crystals. *Proc. R. Soc. Lond. A* **88**, 428–438 (1913).
6. Kirkland, E. J. *Advanced Computing in Electron Microscopy* (Springer, 2010).
7. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
8. Sun, H. COMPASS: an *ab initio* force-field optimized for condensed-phase applications—overview with details on alkane and benzene compounds. *J. Phys. Chem. B* **102**, 7338–7364 (1998).
9. Kobayashi, K., Koshino, M. & Suenaga, K. Atomically resolved images of *I_h* ice single crystals in the solid phase. *Phys. Rev. Lett.* **106**, 206101 (2011).

Author Contributions W.Z., C.W. and A.R.L. wrote this Comment with help from the other authors. W.Z., K.Y., C.W. and T.X. performed the experiment. Y.Z. and S.T.P. performed the calculations. C.W., T.X. and L.S. prepared the graphene samples. All authors discussed the paper and contributed to data analysis.

Competing Financial Interests Declared none.

doi:10.1038/nature16145

Algara-Siller *et al.* reply

REPLYING TO W. Zhou *et al.* *Nature* **528**, <http://dx.doi.org/10.1038/nature16145> (2015)

In the accompanying Comment¹, Zhou *et al.* showed that a NaCl solution captured between graphene sheets leads to the formation of few-layer crystals of NaCl with similar structure and lattice constant as for the ‘square ice’ we described². They suggest that our samples were accidentally contaminated with NaCl or another salt and that the oxygen K-edge in our electron energy-loss spectra (EELS) originates from oxide contaminants on graphene.

We emphasize that at no point were our samples in proximity to NaCl or other salts. All our spectra were obtained in diffraction mode with an effective diameter of 100 nm, not high-resolution imaging mode in which individual crystals may be selected, to decrease the electron dose and allow longer acquisition times. In our EELS analysis, we focused on the energy window in which the oxygen peak was expected; the full energy spectrum comparing regions with and without ice crystals was not acquired in all cases. Unfortunately, this means that we cannot retrospectively prove the absence of NaCl. Nevertheless, following our new simulations of transmission electron microscope (TEM) images of NaCl, the difference in contrast between sodium and chlorine should be visible under our imaging conditions in the case of a mono- or trilayer crystal with a half unit cell. We do not find corresponding differences in contrast in any of our experimental images.

We agree with Zhou *et al.*¹ that our oxygen K spectrum in figure 1b in ref. 2 probably has a contribution from silicon oxide, but we believe this contribution is small. There is disagreement in the literature regarding the peak shape and exact position of the oxygen K-edge for ice. In our paper², we compared the experimental oxygen K-edge (figure 1b, main oxygen K peak at 540 eV) with a simulated

spectrum³ for which the main peak is shifted by approximately 6 eV compared to our experiment. However, other calculations^{4,5} of oxygen K spectra for hexagonal and cubic ice give the oxygen K peak at 540 eV, in agreement with the spectrum in figure 1b in ref. 2. In addition, in our unprocessed oxygen K spectrum, a pre-edge shoulder is seen that is very similar to those in refs 4 and 5. Unfortunately, these weak features are not visible in figure 1b in ref. 2, owing to smoothing of the raw spectrum. Only EELS in high-resolution imaging mode selecting individual crystals (or scanning TEM-EELS) could unambiguously distinguish such features.

In view of the above, further experiments are needed to rule out the contamination hypothesis.

G. Algara-Siller¹, O. Lehtinen¹ & Ute Kaiser¹

¹University of Ulm, 89081 Ulm, Germany.

email: ute.kaiser@uni-ulm.de

1. Zhou, W. *et al.* The observation of square ice in graphene questioned. *Nature* **528**, <http://dx.doi.org/10.1038/nature16145> (2015).
2. Algara-Siller, G. *et al.* Square ice in graphene nanocapillaries. *Nature* **519**, 443–445 (2015).
3. Kobayashi, K., Koshino, M. & Suenaga, K. Atomically resolved images of I_h ice single crystals in the solid phase. *Phys. Rev. Lett.* **106**, 206101 (2011).
4. Prendergast, D. & Galli, G. X-ray absorption spectra of water from first principles calculations. *Phys. Rev. Lett.* **96**, 215502 (2006).
5. Kong, L., Wu, X. & Car, R. Roles of quantum nuclei and inhomogeneous screening in the x-ray absorption spectra of water and ice. *Phys. Rev. B* **86**, 134203 (2012).

doi:10.1038/nature16149

Wang *et al.* reply

REPLYING TO W. Zhou *et al.* *Nature* **528**, <http://dx.doi.org/10.1038/nature16145> (2015)

In the accompanying Comment¹, Zhou *et al.* reproduced our² molecular dynamics (MD) results and pointed out that the simulated 2D ice is slightly rhomboidal, in contrast to the square lattice seen in the transmission electron microscope (TEM) images². We were aware of this disagreement, but did not discuss it in ref. 2 for the following reasons. First, previous MD simulations^{3,4} have reported ‘square ice’, although it remains unclear whether this ice is different to the distorted lattice we found². Second, and more importantly, we were convinced that the simulated, slightly rhomboidal structures should be observed experimentally as square ice.

Indeed, our MD snapshots² (and those presented in ref. 1) show substantial disorder. Each realization is metastable, and the finite temperature is expected to move such defects through the crystal lattice. Our simulations show that this happens on a timescale of tens of nanoseconds for nanometre-sized ice crystals, much longer than the time used by Zhou *et al.*¹, but much shorter than the time needed to obtain experimental images (about 1 s). To simulate this time-averaging effect, we created a number of intermittent states (such as that shown in figure 2d in ref. 2) and superimposed them, keeping the positions of only the edge molecules fixed to simulate the confinement. We found that the slightly rhomboidal lattice averaged out into one that is indistinguishable from a perfect square (not shown in ref. 2).

Finally, perfectly square ice discussed in ref. 2 was subsequently found to be the most stable configuration using first-principle analyses^{5,6}.

Therefore, we maintain that square ice can theoretically occur in hydrophobic nanocapillaries, in agreement with the experiment².

R. R. Nair and I. V. Grigorieva support this Reply, but did not contribute to the part of research that was addressed in the accompanying Comment.

F. C. Wang¹, H. A. Wu¹ & A. K. Geim²

¹University of Science and Technology of China, Hefei, Anhui 230027, China.
email: wuha@ustc.edu.cn

²University of Manchester, Manchester M13 9PL, UK.

1. Zhou, W. *et al.* The observation of square ice in graphene questioned. *Nature* **528**, <http://dx.doi.org/10.1038/nature16145> (2015).
2. Algara-Siller, G. *et al.* Square ice in graphene nanocapillaries. *Nature* **519**, 443–445 (2015).
3. Takaiwa, D., Hatano, I., Koga, K. & Tanaka, H. Phase diagram of water in carbon nanotubes. *Proc. Natl Acad. Sci. USA* **105**, 39–43 (2008).
4. Kolesnikov, A. I. *et al.* Anomalous soft dynamics of water in a nanotube: a revelation of nanoscale confinement. *Phys. Rev. Lett.* **93**, 035503 (2004).
5. Corsetti, F., Matthews, P. & Artacho, E. New ice rules for nanoconfined monolayer ice from first principles. Preprint at <http://arXiv.org/abs/1502.03750> (2015).
6. Chen, J., Schusteritsch, G., Pickard, C. J., Salzmann, C. G. & Michaelides, A. 2D ice from first principles: structures and phase transitions. Preprint at <http://arXiv.org/abs/1508.03743> (2015).

doi:10.1038/nature16146

TUBERCULOSIS

Autophagy is not the answer

The cellular process of autophagy has been proposed to help kill *Mycobacterium tuberculosis*. But although the autophagy gene *Atg5* is key to host immunity, other autophagy genes do not affect the outcome of tuberculosis. [SEE LETTER P.565](#)

SAMUEL M. BEHAR & ERIC H. BAEHRECKE

More people die each year from tuberculosis than from any other infectious disease. The bacterium *Mycobacterium tuberculosis* is transmitted between people by aerosol droplets generated by someone with active lung disease. The immune response prevents disease in most infected people, but the bacteria often persist as an asymptomatic (latent) infection. Many studies have found that the process of autophagy, through which cellular components are broken down and recycled, contributes to the killing of *M. tuberculosis*. In this issue, Kimmey *et al.*¹ (page 565) present evidence supporting the previous finding that the autophagy gene *Atg5* plays a key part in the host response to *M. tuberculosis* infection, but they show that loss of many other autophagy genes does not significantly influence disease outcome. These results suggest that *Atg5* is essential for restricting *M. tuberculosis* growth, but that the conventional autophagy pathway is not.

M. tuberculosis is an intracellular pathogen that infects, persists and replicates in immune cells called macrophages (Fig. 1). These cells are an inhospitable environment for bacteria: engulfment of the bacterium into a membrane-bound compartment known as a phagosome triggers a series of events that leads to the pathogen's destruction in another organelle, the lysosome, in which hydrolytic enzymes break down material targeted for degradation. But *M. tuberculosis* has evolved to survive this engulfment by preventing phagosomal acidification and phagosome and lysosome fusion.

Some people are inherently resistant to infection despite repeated exposure to *M. tuberculosis*. This suggests that cells of the innate immune system, to which macrophages belong, may kill the bacteria. A role for innate immunity in antituberculosis responses is supported by the finding that human macrophages express pattern-recognition receptors on their surface that are triggered by *M. tuberculosis*². The resulting signalling events, in concert with vitamin D signalling, activate macrophages to produce the antibacterial peptide cathelicidin³.

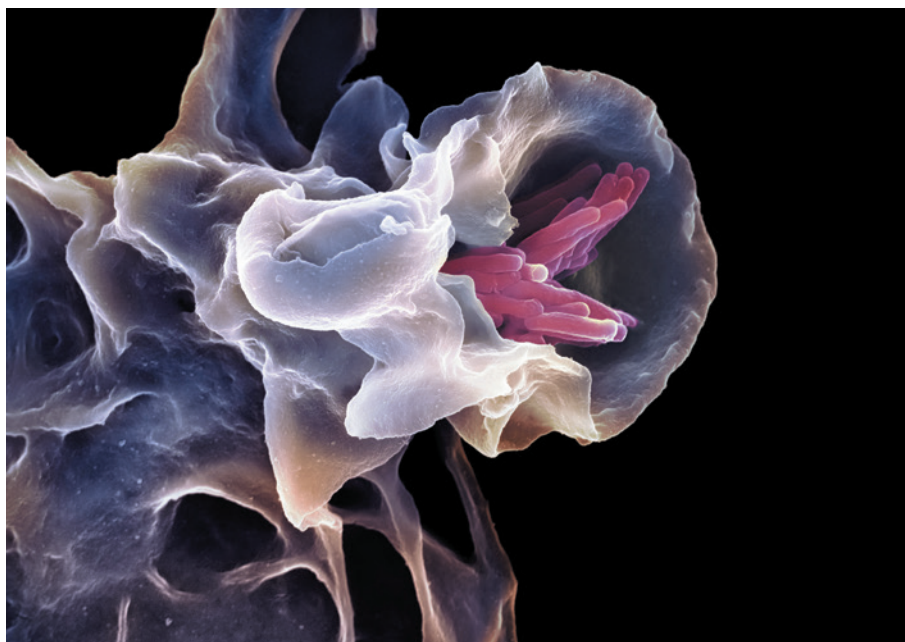


Figure 1 | Defence against *Mycobacterium*. This coloured scanning electron micrograph shows a macrophage cell engulfing *Mycobacterium tuberculosis* cells (pink) by phagocytosis. This process triggers intracellular events leading to bacterial destruction, but some bacteria successfully survive and replicate within the macrophage. It has been suggested that autophagy, a process normally used for degradation and recycling of cellular components, is involved in the immune response against *M. tuberculosis*, but Kimmey *et al.*¹ suggest that the conventional autophagy pathway is not essential for this response *in vivo*.

However, numerous clinical and experimental studies have suggested that T cells — an arm of the adaptive immune system — are needed to synergize with the innate response to control *M. tuberculosis* infection.

T cells produce cytokines and other effector molecules when the cells recognize infected macrophages. One such cytokine is interferon- γ (IFN γ), which is essential for host resistance against mycobacterial infection in animal models and people. IFN γ profoundly alters gene expression in infected macrophages; in mouse macrophages, it stimulates the production of nitric oxide, which is toxic to *M. tuberculosis*. This cytokine also promotes phagosomal acidification and phagosome-lysosome fusion, presumably by overriding the block imposed by *M. tuberculosis*, although the molecular details of this process are unclear. What does seem clear, however, is that the antibacterial activities induced by both innate and

adaptive immune signals require the induction of autophagy^{4,5}. These data raised the possibility that autophagy is the common pathway used by macrophages to restrict the intracellular growth of *M. tuberculosis*.

Autophagy is a fundamental process by which cytoplasmic components, including organelles, are delivered to the lysosome for degradation. This occurs by the formation of a double-membrane vesicle, known as the autophagosome, around the cytoplasmic components being targeted. Alongside its role in cellular homeostasis, autophagy is increasingly found to be involved in host defence. For example, two stimuli known to activate autophagy — starvation and treatment with the drug rapamycin — reduce the viability of *M. tuberculosis* in infected macrophages *in vitro*⁶.

To investigate these observations *in vivo*, researchers turned to mouse

models of tuberculosis. Genetic models that ablate evolutionarily conserved processes such as autophagy are difficult to develop. For instance, the gene *Atg5* encodes part of the protein complex associated with autophagosome membrane formation; mice lacking this gene die soon after birth. However, mouse strains have been developed that lack *Atg5* only in cells of the myeloid lineage, to which macrophages belong. Only a modest loss of bacterial control is seen in these mice following *M. tuberculosis* infection, but large necrotic lung lesions develop, along with increased infiltration of immune cells called neutrophils into the lungs and increased expression of inflammatory cytokines⁷. A separate study in which these mutant mice were infected with a more virulent *M. tuberculosis* strain found a higher level of bacteria in the lungs than in infected normal mice, and faster death from the infection⁸. This study also found necrotic lung lesions and increased levels of inflammatory cytokines in the mice lacking *Atg5* in macrophages. These findings led to the proposal that the ATG5 protein and the autophagy pathway are essential for *M. tuberculosis* control *in vivo*.

Kimmey *et al.* tested this idea using mice that lacked other genetic components required for autophagy in myeloid-lineage cells: *Atg3*, *Atg7*, *Atg12*, *Atg14l* and *Atg16l1*. Surprisingly, the authors did not observe the same characteristics when the mice were infected with *M. tuberculosis*. These data suggest at least two possibilities. One is that ATG5 regulates an autophagy program that does not depend on the other five *Atg* genes tested. Although there is some evidence for the existence of alternative autophagy pathways^{9,10}, these pathways depend on the function of multiple *Atg* genes, and additional data in Kimmey and colleagues' study do not support this conclusion.

An alternative possibility is that ATG5 functions in non-autophagic processes that contribute to *M. tuberculosis* control. *Atg* genes have been implicated in other vesicle-trafficking processes, including endocytosis, protein secretion and LC3-associated phagocytosis^{11–13}. ATG5 interacts with many proteins¹⁴ that could influence pathogenesis, although these proteins are poorly characterized. ATG5 has also been implicated in the regulation of cell death by its association with the proteins Bcl-xL and FADD^{15,16}. Cell death has been linked to inflammation and infection through the recruitment of immune cells, so it is possible that this is the pathway by which ATG5 regulates infection in an autophagy-independent manner.

An intense spotlight has been shining on autophagy as a possible route to designing better vaccines and drugs against *M. tuberculosis* infection. Although Kimmey and colleagues' findings do not mean that the 'reset' button should be hit on such investigations, further studies are needed to determine

whether ATG5 influences *M. tuberculosis* infection through a non-canonical autophagy pathway or through a different cellular process. Once this is established, researchers will be better positioned to develop strategies for the treatment and prevention of tuberculosis. ■

Samuel M. Behar is in the Department of Microbiology and Physiological Systems, and **Eric H. Baehrecke** is in the Department of Molecular, Cell and Cancer Biology, University of Massachusetts Medical School, Worcester, Massachusetts 01655, USA.

e-mails: samuel.behar@umassmed.edu; eric.baehrecke@umassmed.edu

1. Kimmey, J. M. *et al.* *Nature* **528**, 565–569 (2015).
2. Bhatt, K. & Salgame, P. *J. Clin. Immunol.* **27**, 347–362 (2007).
3. Liu, P. T., Stenger, S., Tang, D. H. & Modlin, R. L. *J. Immunol.* **179**, 2060–2063 (2007).
4. Fabri, M. *et al.* *Sci. Transl. Med.* **3**, 104ra102 (2011).

5. Fabri, M. & Modlin, R. L. *Cell Host Microbe* **6**, 201–203 (2009).
6. Gutierrez, M. G. *et al.* *Cell* **119**, 753–766 (2004).
7. Castillo, E. F. *et al.* *Proc. Natl Acad. Sci. USA* **109**, E3168–E3176 (2012).
8. Watson, R. O., Manzanillo, P. S. & Cox, J. S. *Cell* **150**, 803–815 (2012).
9. Chang, T. K. *et al.* *Nature Cell Biol.* **15**, 1067–1078 (2013).
10. Nishida, Y. *et al.* *Nature* **461**, 654–658 (2009).
11. Deretic, V., Jiang, S. & Dupont, N. *Trends Cell Biol.* **22**, 397–406 (2012).
12. Shrivage, B. V., Hill, J. H., Powers, C. M., Wu, L. & Baehrecke, E. H. *Development* **140**, 1321–1329 (2013).
13. Sanjuan, M. A. *et al.* *Nature* **450**, 1253–1257 (2007).
14. Behrends, C., Sowa, M. E., Gygi, S. P. & Harper, J. W. *Nature* **466**, 68–76 (2010).
15. Yousefi, S. *et al.* *Nature Cell Biol.* **8**, 1124–1132 (2006).
16. Young, M. M. *et al.* *J. Biol. Chem.* **287**, 12455–1468 (2012).

This article was published online on 9 December 2015.

COMPUTER TECHNOLOGY

Silicon chips lighten up

Microprocessor communications have received a boost from the integration of electronics and photonics in silicon — a first step towards low power consumption and efficient computing systems. SEE LETTER P.534

LAURENT VIVIE

Ever since the demonstration of the first microprocessor, 'smaller, cheaper, faster' has been the motto of the microelectronics evolution that enables ever-more-densely packed circuits to speed up computer performance. But a bottleneck now exists in terms of speed and power consumption for on-chip data communications; for instance, conventional wires lose energy and reduce the communication speed. A full transition to optical-link technology using photons would overcome these limitations, because photonic devices have no speed constraints and use less energy than conventional electronics¹. Because silicon is the main material for complementary-metal-oxide-semiconductor (CMOS) technology, which is widely used by the electronics industry, it has been the subject of intense research in photonics, giving rise to what has been called the silicon photonics age². Using the same material for electronics and photonics in a single circuit could increase performance and reduce the power consumption of integrated chips. On page 534 of this issue, Sun and co-workers³ report a big advance in such efforts: a microprocessor that communicates using light.

Despite the growing interest in silicon photonics and the development of efficient

integrated devices (circuits) on silicon-on-insulator (SOI) wafers, only a few complete electronic–photonic circuits have been demonstrated. This is because the silicon substrate for photonics is very different from the standard substrates used in electronics — and even slight changes to CMOS technology can degrade the performance of the transistors used in microchips. As such, developing a process to merge electronics and photonics on a single chip is highly challenging^{4,5}.

The first reported strategy for electronic–photonic integration used the 'front-end' approach⁶, in which transistors and photonic devices are placed on the same layer of a silicon chip. The chosen method of fabricating such chips was based on a custom CMOS electronic process on a non-standard SOI substrate, and enables high-speed light propagation on the chips.

But even if this integration solution was reliable for producing efficient on-chip transceivers for data input and output, developing a more complex on-chip system using state-of-the-art electronics would require large investments of money and technological-process development. Furthermore, the main proposed integration solutions would involve a multichip approach² in which the photonic and electronic circuits are fabricated independently using different processes, optimized

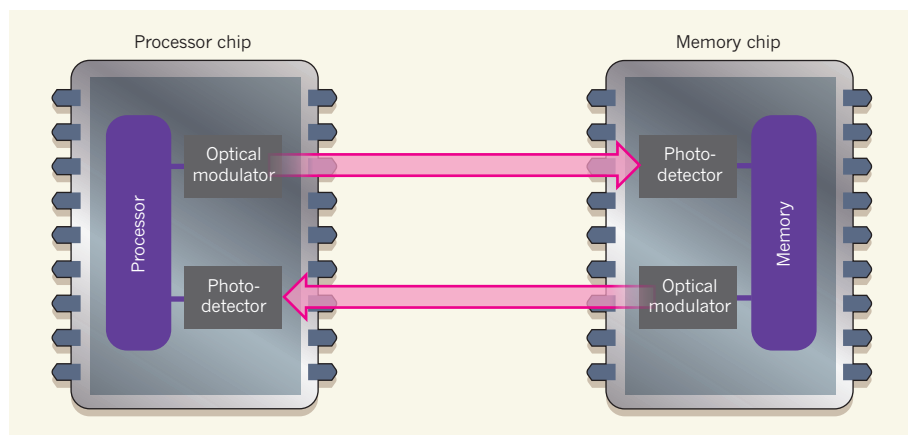


Figure 1 | Processor-memory communication using optical links between two chips. Sun *et al.*³ report a system in which two chips communicate using light. Notably, the system connects electronic components (purple, including the metallic wires) and photonic ones (grey). A silicon optical modulator at each chip converts electronic information into an optical signal to be transmitted to the other chip. The data then travel along an optical fibre (pink) to a detector in the other chip. The detectors transcribe the optical signal into an electronic one. This optical-link technology might increase the performance speed and reduce the power consumption of chips, compared with conventional silicon devices.

for each technology, and brought together through a bonding technique. Such multi-chip integration may be the most economical solution in the short term.

Sun and colleagues' 'zero-change' approach³ challenges this thinking. Based on a commercial SOI CMOS process, it uses existing fabrication steps for electronics, accommodating the photonics without any extra development. This allows all existing electronic designs to be used and combined with photonic components without any additional non-standard processes, which may dramatically increase the efficiency and reliability of the resulting system on a chip (SOC).

The authors report several major advances in the field of microprocessor communication. Their electronic-photonic SOC integrates millions of transistors and hundreds of photonic components to form a microprocessor and memory that communicate with each other using light, at a speed of 2.5 gigabits per second (Fig. 1). The photonic components used to guide, code and detect information are based on a combination of materials that are standard in the electronics industry, including silicon, silicon-germanium (SiGe) and silicon nitride — all of which are implemented in CMOS technology.

The researchers used an external source of light to drive the photonic devices at a wavelength of 1,180 nanometres, with which the light could be confined and channelled efficiently within a waveguide in silicon. To minimize leakage of light from the waveguide into the substrate, the authors selectively and locally remove the substrate under the photonic devices. Each of the two optical links between the microprocessor and the memory includes a compact, silicon micro-ring modulator to code the information at one end, and a SiGe detector driven by both processor and

memory at the other end.

Although the photonic circuit may seem simple, Sun *et al.* optimized it to provide error-free transmission with moderate power consumption. Processors produce heat according to how much they are working, creating large temperature changes over time, which could seriously degrade the performance of optical components. But the authors demonstrate that microprocessor communication is robust in their device under different power conditions (different thermal perturbations), thanks to a feedback loop in the SOC.

Sun and co-workers' result is proof of

concept for the development of a complex electronic-photonic SOC. However, challenges remain before their zero-change approach can be used for the commercial production of such circuits. First, the on-chip optical communication rate of 2.5 gigabits per second is relatively slow compared with the rate achievable by state-of-the-art silicon photonics systems. An increase in the bandwidth of both the optical modulators and the detectors in the team's SOC would increase the performance of the memory-to-processor link.

Second, a multiwavelength optical circuit may be needed in the future to resolve the interconnect bottleneck. Moreover, much larger numbers of photonic devices and functionalities, including switches, filters and delay lines with low power consumption, will one day become necessary to address the future requirements of computing systems. Finally, it would also be beneficial to scale this approach up for use in multicore processors. ■

Laurent Vivien is at the Institut d'Electronique Fondamentale, CNRS, Université Paris-Sud, Université Paris Saclay, 91405 Orsay, France.
e-mail: laurent.vivien@u-psud.fr

1. Miller, D. A. B. *Proc. IEEE* **97**, 1166–1185 (2009).
2. *Handbook of Silicon Photonics* (eds Vivien, L. & Pavesi, L.) (CRC Press, 2013).
3. Sun, C. *et al. Nature* **528**, 534–538 (2015).
4. Fedeli, J.-M. *et al. IEEE J. Sel. Topics Quantum Electron.* **20**, 8201909 (2014).
5. Baehr-Jones, T. *et al. Nature Photon.* **6**, 206–208 (2012).
6. Narasimha, A. *et al. In Optical Fiber Commun. Natl Fiber Optic Eng. Conf.* <http://dx.doi.org/10.1109/OFC.2008.4528356> (2008).

NUTRITION

A personal forecast

Machine learning, applied to complex multidimensional data, is shown to provide personalized dietary recommendations to control blood glucose levels. This is a step towards integrating the gut microbiome into personalized medicine.

ERICA D. SONNENBURG & JUSTIN L. SONNENBURG

Weather forecasters were once the unfortunate subjects of countless jokes. Predicting the weather from multiple interacting meteorological factors that are greatly influenced by underlying geography seemed no better than extracting a forecast from a cloudy crystal ball. The complexity and individuality of the human body pose similar hurdles to making accurate predictions in personalized medicine. But access to huge amounts of data, refinement of mathematical models and enhanced computing power

have transformed predictive meteorology¹, and the same is beginning to apply to human health. Writing in *Cell*, Zeevi *et al.*² approach the complex problem of how an individual's blood glucose concentration will be affected by particular foods, the microorganisms in their gut and other aspects of their physiology, and present a predictive model that enables personalized food recommendations.

Obesity and type 2 diabetes mellitus are sweeping the developed world³. An individual's post-prandial glycaemic response (PPGR), a measure of how much blood glucose levels rise after a meal, is a predictor of risk for developing type 2 diabetes⁴ — the greater the rise,

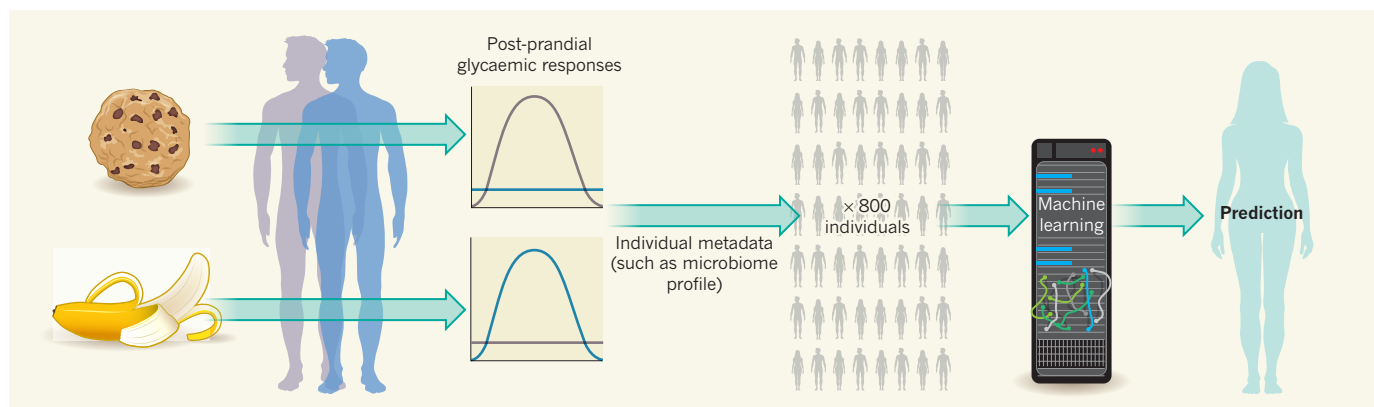


Figure 1 | Machine learning for nutrition advice. Zeevi *et al.*² continuously monitored the blood glucose levels of 800 individuals over the course of a week, which gave an indication of their post-prandial glycaemic responses (PPGRs; a measure of how rapidly blood glucose levels rise after food consumption) to specific foods. They combined this with 137 other measurements from each person, including their body-mass index,

cholesterol levels, diet, activity levels and the composition of their gut microbiome. The data were used to develop a machine-learned algorithm. The authors show that this algorithm can predict PPGRs in people who were not in the cohort used to train the model, and thus can be used to provide dietary recommendations for maintaining PPGRs that are associated with health.

the greater the risk. Because of this link, specific guidelines for how a person can maintain glycaemic control would be extremely useful⁵. Zeevi *et al.* equipped 800 people with subcutaneous probes that measured their blood glucose levels every five minutes over the course of a week (Fig. 1). With the exception of 5,107 standardized meals provided across the group, the participants ate their typical meals and logged detailed dietary records. The contents of the 52,005 meals were then analysed alongside more than 1.5 million glucose measurements.

The data revealed significant interpersonal variability of PPGRs to the identical (standardized) meals and to similar self-reported meals. Furthermore, the foods that induced the highest PPGRs differed greatly between individuals: a banana had a bigger effect than a cookie for one person, but the opposite was true for another. These insights could explain why standard dietary interventions for controlling PPGR are not uniformly effective across a population.

To make sense of the highly personalized glycaemic responses to food, the authors turned to the vast amount of data collected for each individual (Fig. 1). Included in their analyses were physiological characteristics, such as body-mass index; blood markers, such as cholesterol levels; behavioural data gathered from a questionnaire, for example activity level and sleep habits; and profiles of the participants' gut microbiomes (their resident gut microorganisms), including species composition and combined genome sequences. The data immediately revealed that an individual's PPGRs correlate with known risk factors for developing type 2 diabetes, such as body-mass index and systolic blood pressure. However, other, less obvious aspects of the composite medical profile also correlated with PPGRs, including the presence of particular taxa in the microbiome, such as the Enterobacteriaceae,

and particular bacterial genes, such as those involved in chemotactic movement.

The authors then used a 'decision tree' machine-learning method to create an algorithm that would incorporate all these pieces of metadata. This approach proved to be predictive for PPGRs in cross-validation with the cohort of 800 participants — this means that a person's PPGRs could be predicted by an algorithm generated using data from the other 799 participants. The algorithm also predicted PPGRs of an independent cohort of 100 individuals whose data were not used to train the algorithm.

The authors identified several features in the metadata that were associated with an individual's PPGRs. As expected, increased carbohydrate consumption was closely tied to a raised PPGR. The presence of dietary fibre in meals led to an increased PPGR shortly after consumption, but decreased PPGR in the following 24 hours. There were also several features that were predictive of PPGRs that did not relate to meal consumption, including sleep, physical activity and aspects of the microbiome.

Overall, this approach was statistically more accurate at predicting glycaemic response than the current gold standard, which is based on the carbohydrate content of a meal. In a final test, the authors recruited 26 new participants and tailored meal recommendations (such as chicken recommended for one person, but withheld from another) for each participant using either their algorithm or expert interpretation of those individuals' PPGRs to specific meals. The recommendations based on the model improved PPGRs and stability of blood glucose levels similarly to the improvement achieved by the expert recommendations.

Although associations have previously been made between aspects of the gut microbiome and diseases ranging from obesity to autism^{6,7}, the mechanisms underlying such links are

mostly unknown. One of the big advantages of Zeevi and colleagues' approach is that such mechanisms need not be known for it to work. Nevertheless, this study provides a roadmap for generating and testing hypotheses about mechanisms. For example, do the *Akkermansia muciniphila* bacteria, which degrade the glycoprotein mucins that line the gut, and whose presence was found by the authors to correlate with higher PPGRs, causally contribute to this glycaemic response, and if so, how? The authors' large human data set and machine-learning approach provides an excellent launching point for mechanistic studies that are likely to be generalizable and relevant to people.

At this point in time, most microbiome researchers would not want to emulate weather forecasters and be asked to predict an individual's response to diet or medication on the basis of their microbiome profile. However, when combined with a machine-learned algorithm that incorporates additional metrics of host biology, such prediction seems much less daunting. The application of machine-learning methods to endpoints beyond PPGRs, such as progression towards or treatments for autoimmune diseases, cardiovascular disease and cancer, is likely to follow rapidly. In the era of 'big data' science, in which we can measure an enormous number of parameters, harnessing the most-predictive aspects of highly dimensional data will be extremely powerful. Although the previous picture of how the complexity of individual microbiome profiles could inform personalized medicine was cloudy, this study provides the grounds for an optimistic forecast. ■

Erica D. Sonnenburg and Justin L. Sonnenburg are in the Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, California 94305, USA. e-mail: jsonnenburg@stanford.edu

1. Bauer, P., Thorpe, A. & Brunet, G. *Nature* **525**, 47–55 (2015).
2. Zeevi, D. *et al.* *Cell* **163**, 1079–1094 (2015).
3. Chen, L., Magliano, D. J. & Zimmet, P. Z. *Nature Rev.*

- Endocrinol.* **8**, 228–236 (2012).
4. Ludwig, D. S. *J. Am. Med. Assoc.* **287**, 2414–2423 (2002).
5. Willett, W., Manson, J. & Liu, S. *Am. J. Clin. Nutr.*

- 76**, 274S–280S (2002).
6. Wu, H., Tremaroli, V. & Bäckhed, F. *Trends Endocrinol. Metab.* **26**, 758–770 (2015).
7. Hsiao, E. Y. *et al.* *Cell* **155**, 1451–1463 (2013).

MATERIALS SCIENCE

Strength ceiling smashed for light metals

Nanoscale particles have been uniformly dispersed in a magnesium alloy, yielding composites with record-breaking strengths — and raising the prospect of using magnesium as a lightweight metal for structural applications. [SEE LETTER P.539](#)

MARIA TERESA PÉREZ PRADO &
CARMEN M. CEPEDA-JIMÉNEZ

Magnesium has a density two-thirds that of aluminium, one-quarter that of steel and only slightly higher than that of many polymers. It is therefore regarded as a potentially ideal substitute for heavier metals — but magnesium's poorer mechanical behaviour has limited its application. On page 539 of this issue, Chen *et al.*¹ report a method of fabricating magnesium composites that gives the materials the highest specific strength and stiffness of any structural metal. A crucial step in the process is the dispersion of a relatively large volume fraction of

ceramic nanoparticles in the molten metal, overcoming a long-standing challenge in materials technology.

Magnesium is the eighth most common element in Earth's crust and can be extracted from seawater. It is also recycled easily compared with polymers, which makes it environmentally friendly. The first notable commercial use of this metal for civil structural applications was in the Volkswagen Beetle during the 1930s — each car contained 20 kilograms (ref. 2). Bugatti also produced prototypes of a car called the Aerolithe, which had a body made from magnesium (Fig. 1). But the use of magnesium in vehicles was limited throughout the twentieth century because of the high

cost of extracting the metal from its ore, the complexity of its mechanical behaviour, and concerns about its flammability and its susceptibility to corrosion under operational conditions.

Interest in magnesium surged afresh at the turn of this century, motivated by the pressing need to implement environmentally friendly policies in industrial production. The benefits of using lightweight materials have now been amply quantified. For example, a weight reduction of 100 kg for an average car saves about 25 gigajoules of energy and 1,600 kg of carbon dioxide emissions over the car's 10-year lifetime³. A major driver for research in magnesium has been the need to improve its mechanical behaviour dramatically, so that it becomes competitive with widely used heavier materials, such as steel or aluminium alloys.

However, hardening strategies that have led to major improvements in strength for other metals have been less effective for magnesium alloys. For example, the precipitation of a fine dispersion of solid particles of uniform size in aluminium alloys leads to four- to fivefold increases in strength (see ref. 4, for example), because the particles act as obstacles to moving dislocations (crystallographic defects whose movement leads to permanent deformation of materials). Up to now, the most effective precipitation treatments applied to magnesium alloys have barely doubled the alloys' strength⁵.

A major obstacle to further improvement lies in the difficulty of making a uniform dispersion of closely spaced, fine precipitates that effectively hinders the movement of basal dislocations and 'twins' — the deformation modes that are activated in response to the smallest applied stresses. Magnesium alloys often contain a mixture of precipitate phases that have different geometries and size distributions. A more uniformly sized and homogeneous particle distribution could be achieved by optimizing both the alloy composition and the heat treatment. But optimizing both together is extremely complicated, because small changes in alloy composition or in the temperature and duration of a heat treatment can lead to large and unpredictable changes in precipitate distribution.

Another approach to strengthening a metal is to add reinforcing particles of various types, shapes and sizes (usually micrometre-scale or larger). Typical additions to magnesium alloys include ceramic or metallic particles, oxides, borides and, less frequently, carbon or carbon nanotubes. But the resulting materials often have unpredictable mechanical properties, because they are unable to achieve a uniform



Figure 1 | Light vehicles. The low density of magnesium led to its use as a structural material for making cars in the 1930s, such as this Bugatti Aerolithe. But the metal was not widely adopted for this purpose, in part because of concerns about its mechanical properties. Chen *et al.*¹ report a magnesium composite that has much-improved properties for structural applications.

JOE WIECHA

dispersion of particles or good particle–matrix bonding. This limits their applications to niche products.

Powders consisting of nanoscale particles have been proposed to be highly effective reinforcers, and the development of inexpensive methods of producing them in large quantities has attracted a substantial effort to fabricate nanocomposites⁶. However, exploiting the full potential of such materials requires a uniform dispersion of a relatively large volume fraction of individual nanoparticles in the melt of the matrix material⁷. Chen *et al.* have succeeded in fabricating a magnesium–zinc alloy densely populated with individual ceramic nanoparticles (14% volume fraction), and in this way have endowed it with outstanding mechanical behaviour. This is the first time that formation of a nanocomposite has led to such a large increase in strength.

The authors began by dispersing a 1% volume fraction of ceramic nanoparticles in the magnesium alloy in the liquid state, and then increased the concentration of particles by partially evaporating away the metallic alloy in a vacuum furnace. The resulting uniform distribution of nanoparticles (see Fig. 1a and b of ref. 1) is extremely effective in arresting basal slip and twin propagation, leading to an increase in the alloy's yield strength (the stress at which the material starts to deform irreversibly) from around 50 megapascals to around 410 MPa, without impairing plasticity. The nanocomposites also have excellent mechanical stability up to temperatures as high as 400°C.

Chen and colleagues conferred further, extraordinary, strength on the alloy by reducing the size of the grains (small crystals) that make up the bulk metal. The resulting material has a yield strength of 710 MPa, the highest ever reported for polycrystalline magnesium alloys and their composites.

The authors' preparation method has been validated at the laboratory scale, and seems to be particularly suited to fabricating small components made from metals that have melting points similar to, or lower than, that of magnesium (aluminium or zinc, for example). However, further work would be needed to optimize the processing parameters for other metals.

It remains to be seen whether the method could be feasible and environmentally friendly on an industrial scale. Potential problems in scaling up the process might include: the amount and cost of the energy needed; elimination of toxic residues from the evaporated matrix material; and maintenance of the equipment used. In addition, fabricating large amounts of homogeneous nanocomposite could be extremely difficult, because gradient distributions of particles are likely to develop during processing.

But there is no doubt that Chen and colleagues' work constitutes a milestone in our quest to design lighter, stronger materials, and

opens up fresh avenues for the development of metals with unprecedented properties. For example, by choosing appropriate particles and optimizing their spatial distribution, it might be possible to make nanocomposites that have enhanced magnetic and electrical properties compared with existing materials. ■

Maria Teresa Pérez Prado and Carmen M. Cepeda-Jiménez are at the IMDEA Materials Institute, 28906 Getafe, Madrid, Spain.
e-mails: teresa.perez.prado@imdea.org;
carmen.cepeda@imdea.org

MICROBIOLOGY

A division of labour combined

The discovery of microorganisms that can oxidize ammonia all the way to nitrate refutes the century-old paradigm that this nitrification process requires the activity of two types of microbe. SEE ARTICLE P.504 & LETTER P.555

MARCEL M. M. KUYPERS

Bioavailable nitrogen is essential for all organisms and is the main limiting nutrient for life on our planet. This nitrogen enters the environment as ammonia produced by microbial or industrial fixation of nitrogen gas. It is lost, again as nitrogen gas, when microorganisms respire oxidized nitrogenous compounds, such as nitrate and nitrite, instead of oxygen. The process of nitrification — the oxidation of ammonia to nitrate by way of nitrite — links the gain and loss of bioavailable nitrogen and thus plays a central part in the nitrogen cycle. Since its discovery in 1890 (ref. 1), nitrification has been thought to be performed as a 'labour union', with distinct microorganisms carrying out the two steps. In this issue, Daims *et al.*² (page 504) and van Kessel *et al.*³ (page 555) independently show that microorganisms from the genus *Nitrospira* can conduct complete nitrification on their own.

The nitrifying unions known until now involved bacteria or archaea that oxidize ammonia to nitrite, and then different bacteria that oxidize nitrite to nitrate (Fig. 1). It had, however, been predicted that a single organism should be able to carry out both steps of this process, on the basis that full nitrification is energetically highly favourable⁴. The term comammox, for 'complete ammonia oxidation', was coined for this hypothetical process. But for more than a century, only partnerships of ammonia oxidizers and nitrite oxidizers were detected in microbial communities that performed nitrification.

1. Chen, L.-Y. *et al.* *Nature* **528**, 539–543 (2015).
2. Mang, T., Bobzin, K. & Bartels, T. *Industrial Tribology* (Wiley-VCH, 2011).
3. Helms, H. & Lambrecht, U. *Int. J. Life Cycle Assess.* **12** (spec. issue), 58–64 (2007).
4. Deschamps, A. & Brechet, Y. *Acta Mater.* **47**, 293–305 (1998).
5. Nie, J.-F. *Metallurg. Mater. Trans. A* **43**, 3891–3939 (2012).
6. Sillekens, W. *et al.* *Metallurg. Mater. Trans. A* **45**, 3349–3361 (2014).
7. Ferguson, J. B., Sheykh-Jaberi, F., Kim, C.-S., Rohatgi, P. K. & Cho, K. *Mater. Sci. Eng. A* **558**, 193–204 (2012).

At first, nothing seemed to be different about the communities that were enriched by Daims *et al.* from a deep oil well and by van Kessel *et al.* from an aquaculture system. The microorganisms from both samples grew on ammonia and produced nitrate, just like typical labour unions of ammonia and nitrite oxidizers. However, no known ammonia oxidizers were present in the cultures, whereas species of the bacterial genus *Nitrospira* were abundant. The *Nitrospira* genus belongs to an ancient phylum of bacteria, the Nitrospirae, members of which were previously thought to carry out oxidation only of nitrite to nitrate⁵.

By reconstructing the genomes of the enriched *Nitrospira* species, the researchers found that the organisms had genes related to those used for ammonia oxidation in other organisms. Daims and co-workers further show that the *Nitrospira* species they report expressed these genes when grown on ammonia. And when van Kessel and co-workers provided their *Nitrospira* species with a fluorescently labelled analogue of ammonia, which binds to the organism's ammonia-oxidizing enzyme, they observed fluorescently labelled *Nitrospira* cells, indicating that the microbes oxidized ammonia to nitrite and then nitrate.

Both research groups also used the distinctive ammonia-oxidation gene sequences in these organisms to search for comammox bacteria in other environments. They found the gene sequences to be widespread in both man-made environments and a variety of natural terrestrial ecosystems. Surprisingly, however,

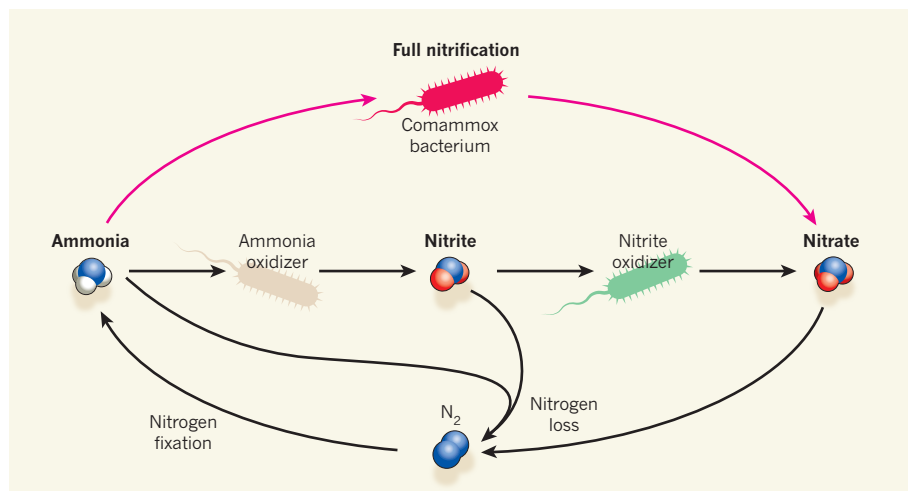


Figure 1 | The nitrogen cycle revised. The global nitrogen cycle begins when ammonia is produced through the fixation of nitrogen gas (N₂) by microorganisms or industrial processes. For N₂ then to be returned to the atmosphere by microbial respiration, the ammonia must undergo nitrification. This process is known to occur in a stepwise fashion when microorganisms oxidize ammonia to nitrite, and other microorganisms oxidize nitrite to nitrate. Daims *et al.*² and van Kessel *et al.*³ now show that 'comammox' bacteria of the *Nitrospira* genus can conduct both steps of nitrification.

the gene sequences were not found in marine waters. Oceans cover about two-thirds of our planet and are sites of intense nitrogen cycling, in which ammonia-oxidizing archaea^{6,7} and nitrite-oxidizing bacteria play a major part. Yet most of these waters are characterized by low ammonia availability, which should theoretically⁴ favour comammox organisms that do not have to share the energy gained from nitrification with a partner organism.

So why were no comammox genes found in ocean waters? Do factors such as salinity stress, the capacity to use organic nitrogen instead of ammonia for nitrification, requirements for specific micronutrients or viral infections lead to comammox organisms being at a competitive disadvantage in the ocean? Or do comammox organisms that have different ammonia-oxidizing genes inhabit the marine realm? It is likely that comammox organisms thrive in the oceans, and environmental microbiologists are in for an exciting time searching for them.

Of the many questions still to be answered, one pertains to microbial production of nitrous oxide. Labour unions of nitrifying organisms are assumed to be one of the main sources of atmospheric nitrous oxide, a potent greenhouse gas and a major contributor to ozone destruction⁸, yet it seems that only the ammonia-oxidizing partners produce the gas. There is no evidence that comammox organisms produce nitrous oxide, but they probably do, because their ammonia-oxidizing pathway is similar to that of classic ammonia-oxidizing bacteria. The identification and cultivation of these organisms by Daims *et al.* and van Kessel *et al.* will inspire the exploration of this and other questions about the role of fully nitrifying organisms in the global nitrogen cycle.

During the past two decades, the discoveries

of several classes of microorganism have profoundly changed our view of the nitrogen cycle. These include anaerobic ammonia-oxidizing bacteria⁹, ammonia-oxidizing archaea¹⁰, methane-oxidizing organisms that generate oxygen from toxic nitric oxide¹¹, symbiotic heterotrophic nitrogen-fixing

cyanobacteria¹² and phototrophic nitrite oxidizers¹³. Comammox organisms are another key addition, and their discovery is proof that if a process is energetically feasible, it will be performed by a microorganism or a microbial labour union somewhere. ■

Marcel M. M. Kuypers is in the Biogeochemistry Department, Max Planck Institute for Marine Microbiology, 28359 Bremen, Germany.
e-mail: mkuypers@mpi-bremen.de

1. Winogradsky, S. *Ann. Inst. Pasteur* **4**, 213–231 (1890).
2. Daims, H. *et al. Nature* **528**, 504–509 (2015).
3. van Kessel, M. A. H. J. *et al. Nature* **528**, 555–559 (2015).
4. Costa, E., Pérez, J. & Kreft, J.-U. *Trends Microbiol.* **14**, 213–219 (2006).
5. Lüscher, S. *et al. Proc. Natl Acad. Sci. USA* **107**, 13479–13484 (2010).
6. Wuchter, C. *et al. Proc. Natl Acad. Sci. USA* **103**, 12317–12322 (2006).
7. Francis, C. A., Roberts, K. J., Beman, J. M., Santoro, A. E. & Oakley, B. B. *Proc. Natl Acad. Sci. USA* **102**, 14683–14688 (2005).
8. Ravishankara, A. R., Daniel, J. S. & Portmann, R. W. *Science* **326**, 123–125 (2009).
9. Strous, M. *et al. Nature* **400**, 446–449 (1999).
10. Könneke, M. *et al. Nature* **437**, 543–546 (2005).
11. Ettwig, K. F. *et al. Nature* **464**, 543–548 (2010).
12. Thompson, A. W. *et al. Science* **337**, 1546–1550 (2012).
13. Griffin, B. M., Schott, J. & Schink, B. *Science* **316**, 1870 (2007).

REGENERATIVE BIOLOGY

Innate immunity repairs gut lining

It emerges that innate immune cells called group 3 innate lymphoid cells signal directly to intestinal stem cells to promote the replacement of damaged epithelial cells lining the gut. SEE LETTER P.560

KONRAD GRONKE & ANDREAS DIEFENBACH

The epithelial-cell layer that lines the intestine acts as a protective barrier against a plethora of microbes and toxic nutrients. As such, it often needs to be rapidly repaired — a process that is initiated by intestinal stem cells (ISCs), which reside in specialized niches at the bottom of small pits called crypts in the intestinal wall¹. In steady-state conditions, to compensate for the normal continual loss of epithelial cells, ISCs divide once every 24 hours¹ to generate progeny that differentiate into all the epithelial-cell types found in the intestine. In response to damage, changes in the behaviour of ISCs are typically thought to be directed by mediator signals released from surrounding epithelial

cells in the stem-cell niche, but on page 560 of this issue, Lindemans *et al.*² reveal that the true picture is much broader.

Two cell types have been implicated in promoting ISC differentiation — secretory epithelial cells called Paneth cells³, which are interspersed throughout the niche, and surrounding connective-tissue cells such as stromal cells, which reside adjacent to the niche. Both of these cell types provide ISCs with essential growth and differentiation factors⁴, but are they the only cell types that regulate stem-cell behaviour?

In 1996, it was found that several thousand follicles containing group 3 innate lymphoid cells (ILC3s) directly underlie intestinal crypts, and are therefore in close proximity to ISCs and their niches (reviewed in ref. 5). ILC3s are

part of a subgroup of haematopoietic (blood) cells called lymphocytes that are involved in innate immunity⁶, and are known to be involved in tissue protection and in fortifying barrier surfaces^{7,8}. However, the crosstalk between follicle-resident ILC3s, ISCs and niche cells has not been explored.

ILC3s produce interleukin-22 (IL-22), a soluble protein that signals to non-haematopoietic cells⁹, and studies indicate that IL-22 can instruct epithelial repair^{7,8}. Indeed, the group that performed the current study previously demonstrated¹⁰ that IL-22 protects ISCs against damage in graft-versus-host disease (GvHD) — a frequent complication of haematopoietic-stem-cell transplantation (HSCT), which is a clinical procedure used to treat haematopoietic tumours. During GvHD, immune cells called T lymphocytes from the donor attack the recipient's tissues, causing severe inflammation. GvHD manifests most frequently in the intestinal epithelium and the skin, and can cause severe organ damage and death¹¹.

Together, then, evidence indicates that the role of the immune system might be more extensive than generally thought, encompassing organ maintenance and epithelial repair in the intestine. But whether ILC3-derived IL-22 can directly control the behaviour of stem cells, or if IL-22 might act on niche cells, has remained unclear. To address this question, Lindemans and colleagues made use of 'mini-guts' grown *in vitro* from single stem cells³. These intestinal organoids faithfully recapitulate the main features of normal gut epithelium — they comprise many crypts that contain ISCs and all ISC-derived cell types, and each crypt feeds into a central lumen lined by mature, absorptive epithelial cells. The authors grew these mini-guts from purified ISCs in the presence or absence of IL-22-producing ILC3s or IL-22. IL-22 substantially increased organoid size and crypt budding, but did not affect the numbers of organoids generated.

On which cell type does IL-22 act? Paneth cells are known to provide niche support for stem cells³. However, Lindemans *et al.* showed that ISCs, but not Paneth cells, express high levels of the IL-22 receptor protein, suggesting that IL-22 acts directly on ISCs. The authors found that the effects of IL-22 were maintained in organoids engineered to lack Paneth cells, suggesting that Paneth-cell-derived signals do not coordinate the effects of IL-22 on ISCs. Furthermore, when mice were infused with donor T lymphocytes to induce GvHD, treatment with IL-22 ameliorated the disease and prevented the loss of ISCs that accompanies intestinal GvHD. Thus, IL-22 directly affects regeneration of the ISC pool after GvHD-mediated damage, augmenting ISC-mediated epithelial repair independently of Paneth-cell-derived signals (Fig. 1).

This study extends our understanding of the niche, identifying ILC populations as a

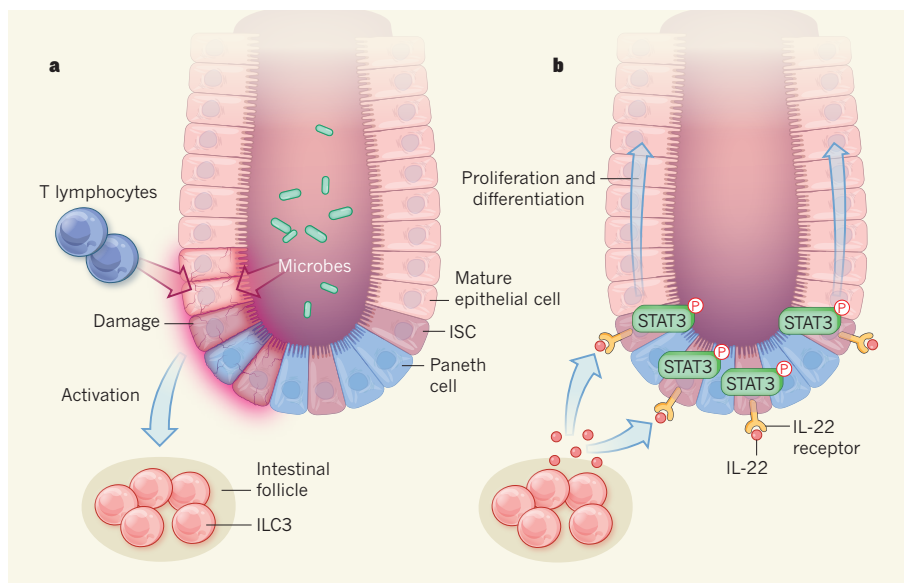


Figure 1 | Damage response. **a**, The intestine is lined by a layer of mature epithelial cells, which are derived from intestinal stem cells (ISCs) that are interspersed with secretory Paneth cells in niches at the bottom of small intestinal pits called crypts. Paneth cells provide the signals that enable stem cells to proliferate and replenish the epithelium under normal conditions. Lindemans *et al.*² report that a different factor replenishes the epithelium following damage caused, for instance, by foreign immune cells called T lymphocytes that cause graft-versus-host disease or by harmful microbes. Damage leads to activation of group 3 innate lymphoid cells (ILC3s), which are located in follicles adjacent to the ISC niche. **b**, On sensing damage, ILC3s release the protein IL-22, which binds to the IL-22 receptor protein on ISCs. This triggers phosphorylation (P) of the protein STAT3, and enhances ISC proliferation, replenishing the damaged epithelial-cell populations.

previously unappreciated checkpoint in tissue repair — a contrast to current models, which hold that only stromal cells and the specialized progeny of stem cells provide niche signals⁴. Although a role for stromal cells in ISC support should not be entirely discarded, the discovery of the direct effect of IL-22 on stem cells will change how researchers think about the stem-cell niche. In the past, haematopoietic cells have been considered to be beneficiaries of niche-derived growth factors, but now it becomes clear that they themselves can also provide supporting factors that maintain the fitness of epithelial-cell lineages.

Lindemans and colleagues' data will open up avenues for future research. ILC3s seem to either directly or indirectly 'sense' epithelial damage and respond with an increased release of IL-22, but the molecular machinery used by ILC3s to sense such damage remains to be defined. Furthermore, it is unclear which IL-22-dependent molecular circuitry coordinates stem-cell behaviour. The authors found that IL-22 signalling leads to phosphorylation of the signalling protein STAT3, thus driving STAT3 signalling in ISCs (Fig. 1). However, ISCs lacking STAT3 lost their stem-cell potential and so could not generate organoids, preventing analysis of the downstream effectors of this signalling pathway. Future studies using genome-wide transcriptional profiling of ISCs deprived of IL-22 signals might reveal the relevant molecular targets.

It is exciting to consider that application of IL-22 might help to ameliorate GvHD — one

of the most serious and limiting effects of HSCT. Interestingly, the unleashing of donor-derived T lymphocytes on host tissues is not only a complication of HSCT, but also a desired effect, because the cells' ability to attack haematopoietic tumours is vital for eradication of the disease¹¹. Clinicians have long tried to balance such beneficial effects of HSCT with the risk of developing GvHD. The finding that IL-22 can prevent damage to ISCs may help to minimize collateral damage and maximize the efficacy of HSCT. ■

Konrad Gronke and Andreas Diefenbach are at the Research Centre for Immunology, University of Mainz Medical Centre, and at the Institute of Medical Microbiology and Hygiene, 55131 Mainz, Germany. K.G. is also at the Max-Planck-Institute of Immunobiology and Epigenetics, Freiburg, Germany. e-mail: diefenbach@uni-mainz.de

1. Clevers, H. *Cell* **154**, 274–284 (2013).
2. Lindemans, C. A. *et al.* *Nature* **528**, 560–564 (2015).
3. Sato, T. *et al.* *Nature* **469**, 415–418 (2011).
4. Morrison, S. J. & Spradling, A. C. *Cell* **132**, 598–611 (2008).
5. Eberl, G. *Nature Rev. Immunol.* **5**, 413–420 (2005).
6. Diefenbach, A., Colonna, M. & Koyasu, S. *Immunity* **41**, 354–365 (2014).
7. Pickert, G. *et al.* *J. Exp. Med.* **206**, 1465–1472 (2009).
8. Aparicio-Domingo, P. *et al.* *J. Exp. Med.* **212**, 1783–1791 (2015).
9. Wolk, K. *et al.* *Immunity* **21**, 241–254 (2004).
10. Hanash, A. M. *et al.* *Immunity* **37**, 339–350 (2012).
11. Blazar, B. R., Murphy, W. J. & Abedi, M. *Nature Rev. Immunol.* **12**, 443–458 (2012).

This article was published online on 9 December 2015.

2015

EDITORS' CHOICE

Extracts from selected
News & Views articles published
this year.

QUANTUM PHYSICS

DEATH BY EXPERIMENT FOR LOCAL REALISM

Howard Wiseman (*Nature* 526, 649–650; 2015)

The world is made up of real stuff, existing in space and changing only through local interactions. Quantum mechanics implies that this intuitive local-realism hypothesis is false. However, no definitive experiment has disproved it — until now. Hensen *et al.* report the first violation of a constraint, called a Bell inequality, under conditions that prevent alternative explanations of the experimental data. A Bell inequality is a mathematical relationship regarding the statistics of measurements obtained by two or more parties, and also involving the measurement settings. Suppose that the parties are in well-separated laboratories, and that the measurement settings are chosen and implemented, and the outcomes obtained, in a sufficiently short time that the only way the choice of setting by any party could affect the outcome of any other party would be through faster-than-light influence. Then, by definition, all Bell inequalities will be satisfied by all local-realistic theories. An experiment violating a Bell inequality implies that either locality or realism is false. Hensen and colleagues' experiment hammers the final nail in the coffin of local realism.

Original research: *Nature* 526, 682–686 (2015).

STEM CELLS

ASYMMETRIC REJUVENATION

Anu Suomalainen (*Nature* 521, 296–298; 2015)

The difference between the daughter cells of an asymmetric stem-cell division is not subtle. One daughter inherits the mother's immortality, whereas the other must leave the cosy stem-cell home and commit to differentiating into a specialized cell type. Writing in *Science*, Katajisto *et al.* observed that organelles called mitochondria showed differential segregation during asymmetric stem-cell division, the stem-cell daughter receiving most of the newly synthesized mitochondria. Mitochondria use oxygen to burn fats, sugars and amino acids, generating side products called reactive oxygen species (ROS) — potent signalling molecules that, if produced in excess, can be damaging. Fully functional mitochondrial proteins minimize ROS production. It is therefore no surprise that stem cells treasure prime fitness in this organelle. Alternatively, perhaps the committed daughter cell requires old mitochondria. An increase in ROS is associated with differentiation; the asymmetric apportioning of mitochondria could therefore provide the ROS boost required to initiate a differentiation program.

Original research: *Science* 348, 340–343 (2015).



NASA/NOAA

PLANETARY SCIENCE

THE MOON'S TILT FOR GOLD

Robin Canup (*Nature* 527, 455–456; 2015)

A giant impact with Earth is thought to have created an Earth-orbiting disk of debris that formed the Moon. A Moon that assembled from 'inelastic' collisions between such debris would orbit approximately in Earth's equatorial plane. Yet the Moon's current orbit implies that its initial orbit was substantially inclined relative to Earth's Equator. Pahlevan and Morbidelli use computational methods to consider the effects of large background objects, such as those that may have delivered the last roughly 1% of Earth's mass, on the Moon's early orbit. An object approaching the Moon from a random direction may increase or decrease the Moon's orbital tilt. The authors' results show a high likelihood that such random scattering events can cumulatively produce the necessary tilt in the Moon's orbit, as long as the number of objects that deliver the final approximately 1% of Earth's mass is small (fewer than 5) and the rate of early tidal expansion of the Moon's orbit is sufficiently rapid.

Original research: *Nature* 527, 492–494 (2015).

COMPUTATIONAL BIOLOGY

HOW TO CATCH RARE CELL TYPES

Lu Wen & Fuchou Tang (*Nature* 525, 197–198; 2015)

How many cell types are there in the human body? Scientists are now addressing this question in a systematic and non-biased way. Grün *et al.* used single-cell sequencing of the transcriptome (the complete collection of RNA molecules in a cell) to analyse 238 cells obtained from mouse 'mini guts' grown *in vitro*. Standard clustering algorithms could not distinguish subgroups within the rare secretory-cell lineage, which was represented by only 20 of the cells. To get around this limitation, Grün *et al.* developed RaceID, a clever algorithm that assumes that a given cell type is likely to strongly express a certain number of cell-type-specific 'outlier' genes. Such genes can be identified if care is taken to exclude technical and biological noise. Using RaceID, the authors identified new secretory-cell subtypes, and validated them *in vivo*. Through the unremitting efforts of Grün *et al.* and others, in the near future we may be able to chart a complete cell-lineage map of the human body.

Original research: *Nature* 525, 251–255 (2015).

365 DAYS: the year in science

CLIMATE SCIENCE

UNBURNABLE FOSSIL-FUEL RESERVES

Michael Jakob & Jérôme Hilaire (*Nature* 517, 150–152; 2015)

The implementation of ambitious climate policies would lead to large proportions of fossil-fuel reserves remaining unexploited. McGlade and Ekins comprehensively quantify the regional distribution of reserves that should not be burned between 2010 and 2050, by modelling a broad range of scenarios based on least-cost climate policies. About 80%, 50% and 30% of coal, gas and oil reserves, respectively, would need to remain below Earth's surface if the world is to limit an increase in global mean temperature to 2°C. The authors' results clearly highlight the distributional challenge of climate policy: imposing a limit on the use of fossil fuels transfers economic benefits (known as rents) from resource owners to those who obtain the right to use the remaining burnable reserves. Hence, successful climate policy will crucially hinge on the question of whether this 'climate rent' can be shared in an equitable way that also ensures resource owners are compensated for their losses.

Original research: *Nature* 517, 187–190 (2015).

MALARIA

FIFTEEN YEARS OF INNOVATIONS

Janet Hemingway (*Nature* 526, 198–199; 2015)

A child still dies every minute from malaria. To drive down this burden further, we need to attribute the contributions of different interventions and use this information to optimize our efforts. Bhatt *et al.* used authoritative, data-driven models to estimate the relative impact that drugs and mosquito-control strategies have had across Africa since 2000. They found that 663 million clinical cases of malaria were averted between 2000 and 2015, and that 68% of these were due to insecticide-treated bednets. Although this massive improvement in malaria control should be applauded, the study provides a timely warning against complacency. The interventions are increasingly threatened by mosquito resistance to insecticides or parasite resistance to drugs. But if we can overcome

hurdles in the development and roll-out of new agents, develop an effective vaccine to reduce transmission and optimally deploy these interventions, then no child need die from malaria.

Original research: *Nature* 526, 207–211 (2015).

PALAEONTOLOGY

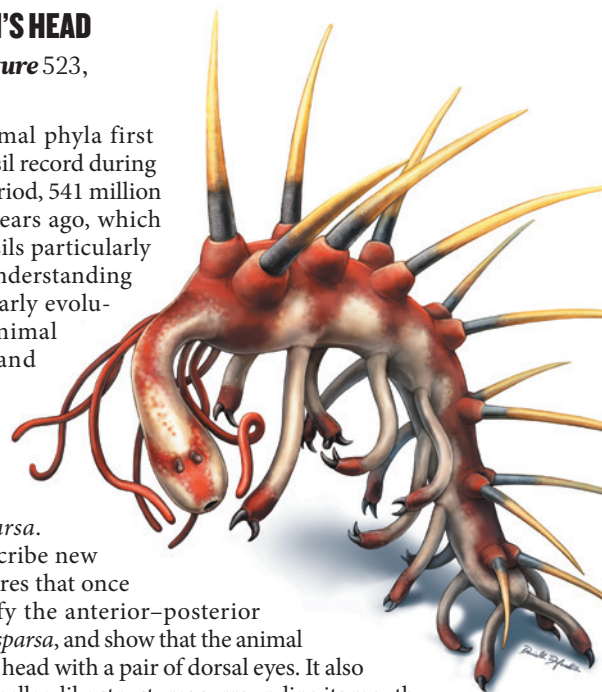
HALLUCIGENIA'S HEAD

Xiaoya Ma (*Nature* 523, 38–39; 2015)

Most major animal phyla first appear in the fossil record during the Cambrian period, 541 million to 485 million years ago, which makes these fossils particularly important for understanding the origin and early evolution of major animal groups. Smith and Caron redescribe one of the most celebrated Cambrian fossil animals, *Hallucigenia sparsa*.

The authors describe new anatomical features that once and for all clarify the anterior–posterior orientation of *H. sparsa*, and show that the animal had an elongated head with a pair of dorsal eyes. It also had hardened, lamellae-like structures surrounding its mouth opening (circumoral elements), and the front part of its foregut (its pharynx) was lined with teeth. These morphological features are suggested to be two of the few characters uniting all groups within the Ecdysozoa — the richest animal group.

Original research: *Nature* 523, 75–78 (2015).



ORGANIC CHEMISTRY

A CURE FOR CATALYST POISONING

Marcus E. Farmer & Phil S. Baran (*Nature* 524, 164–165; 2015)

Capsules or pills remain the most common formulation mode for drugs — without them, pharmacists would have to carefully weigh out and dispense freshly prepared powders of drug substances to patients. Yet research chemists still have to do this for each compound used in their reactions. This is especially problematic when using reagents and catalysts that are sensitive to atmospheric water vapour, oxygen or carbon dioxide. Buchwald and colleagues describe an ingenious solution to this problem: a technique that places reagents and catalysts in a paraffin-wax capsule. Because paraffin wax is generally unreactive, the capsule can simply be added directly to reaction mixtures using common laboratory procedures. The capsule melts on heating, releasing its contents, and the molten wax does not interfere with the desired chemical reaction. If many catalysts and reagents become readily available as capsules, the influence of this approach will probably be seen in the pharmaceutical, agricultural and materials industries.

Original research: *Nature* 524, 208–211 (2015).



DANIELLE DUFAULT

NYANI QUARMYNE/PANOS

Network-analysis-guided synthesis of weisaconitine D and liljestrandinine

C. J. Marth^{1*†}, G. M. Gallego^{1*†}, J. C. Lee^{1†}, T. P. Lebold^{1†}, S. Kulyk^{1†}, K. G. M. Kou¹, J. Qin², R. Lilien² & R. Sarpong¹

General strategies for the chemical synthesis of organic compounds, especially of architecturally complex natural products, are not easily identified. Here we present a method to establish a strategy for such syntheses, which uses network analysis. This approach has led to the identification of a versatile synthetic intermediate that facilitated syntheses of the diterpenoid alkaloids weisaconitine D and liljestrandinine, and the core of gomandonine. We also developed a web-based graphing program that allows network analysis to be easily performed on molecules with complex frameworks. The diterpenoid alkaloids comprise some of the most architecturally complex and functional-group-dense secondary metabolites isolated. Consequently, they present a substantial challenge for chemical synthesis. The synthesis approach described here is a notable departure from other single-target-focused strategies adopted for the syntheses of related structures. Specifically, it affords not only the targeted natural products, but also intermediates and derivatives in the three families of diterpenoid alkaloids (C-18, C-19 and C-20), and so provides a unified synthetic strategy for these natural products. This work validates the utility of network analysis as a starting point for identifying strategies for the syntheses of architecturally complex secondary metabolites.

Chemical synthesis is fundamental to the preparation of small-molecule active pharmaceutical ingredients^{1–4}. Advances in the field of chemical synthesis continue to be marked by the methods and strategies for the preparation of complex natural products, which, more effectively than any other exercise, expose challenges that still exist in the field^{5,6}. Over the last half century, natural product synthesis has continued to be driven by three general motivations: (1) to achieve the practical synthesis of highly complex structures for which a synthesis plan is not readily apparent; (2) to highlight the power, and identify the scope and limitations, of a newly developed synthesis method; and (3) to facilitate exploration of biological function of the synthetically prepared molecules (and their derivatives). Although the second and third motivations have received considerable attention (especially over the last two decades), the first motivation, which has historically served to advance the field, has waned as the notion that any desired molecule can be prepared given enough resources and time has prevailed^{7–9}. Yet, efficient and versatile syntheses of many complex molecules still have not been realized. This is especially true for molecules that feature polycyclic, highly caged frameworks for which effective strategic solutions are not immediately obvious. For these architecturally complex skeletons (for example, aconitine, **1**, Fig. 1a), the biosynthetic transformations that lead to these secondary metabolites in nature are often not fully vetted, are low yielding, or cannot be efficiently reproduced in the laboratory^{10,11}. Therefore, *de novo* strategic approaches for their chemical syntheses are required¹².

Here, we demonstrate that for a subset of topologically complex and functional-group-dense secondary metabolites in the diterpenoid alkaloid family (representative of the aconitine structural type; >700 members), the iterative application of network analysis at the initial stages of synthetic planning yields a unified strategy for their synthesis. This type of analysis has proved to be highly enabling, by identifying a strategy that is a notable departure from previously

established synthesis strategies for related alkaloids. The network analysis approach¹³ involves 'strategic bond disconnections' of bridged polycycles. Despite the emergence of other philosophies, guidelines and methods for synthesis, network analysis remains immutable. Total syntheses of weisaconitine D (**2**; a C-18 alkaloid) and liljestrandinine (**3**; C-19), and the preparation of the skeleton of natural products in the denudatine-type diterpenoid alkaloids (for example, gomandonine, **4**; C-20) reported here illustrate the power of this type of analysis.

The diterpenoid alkaloids (including weisaconitine D and liljestrandinine) have also gained in prominence as small-molecule ligands for voltage-gated Na⁺ and K⁺ ion channels¹⁴. In some cases, these small molecules may be isoform-specific in their interactions with ion channels (presumably binding at the aconitine binding site) and therefore hold potential as the basis for new therapeutics to address myriad channelopathies^{15,16}; for example, the Na⁺ channel blocker lappaconitine (allapinin; **5**) is already administered as a non-narcotic analgesic drug¹⁷. However, to better identify the salient features of these molecules that lead to desirable medicinal properties, versatile *de novo* syntheses are required, because they facilitate the synthesis of analogues featuring deep-seated skeletal changes that might not be otherwise efficiently accessed (for example, by a biomimetic pathway or semi-synthesis).

Network analysis as a starting point in retrosynthesis

The application of network analysis to the diterpenoid alkaloids is illustrated in our retrosynthesis of the C-18 diterpenoid alkaloid weisaconitine D (Fig. 1b). The aim of this analysis is to minimize, in the retrosynthetic direction, the number of bridged rings, which, in addition to the density of stereochemically disposed functional groups, heightens the complexity of these molecules. Targeting the maximally bridged ring (highlighted in red for perspective IV of **2**; see box in Fig. 1b) possessing five bridgehead atoms (highlighted in purple), for disconnection leads back to **6**, to which a bicyclization/cycloaddition

¹Department of Chemistry, University of California, Berkeley, California 94720, USA. ²Cadre Research Labs, Chicago, Illinois 60654, USA. [†]Present addresses: The University of Chicago Law School, 1111 East 60th Street, Chicago, Illinois 60637, USA (C.J.M.); Department of Chemistry, Pfizer Pharmaceuticals, La Jolla Laboratories, 10770 Science Center Drive, La Jolla, California 92121, USA (G.M.G.); Worldwide Medicinal Chemistry, Groton Laboratories, Pfizer Inc. Eastern Point Road, Groton, Connecticut 06340, USA (J.C.L.); Janssen Research & Development, LLC, 3210 Merryfield Row, San Diego, California, 92121-1126, USA (T.P.L.); Theravance Biopharma US Inc., 901 Gateway Boulevard, South San Francisco, California 94080, USA (S.K.).

*These authors contributed equally to this work.

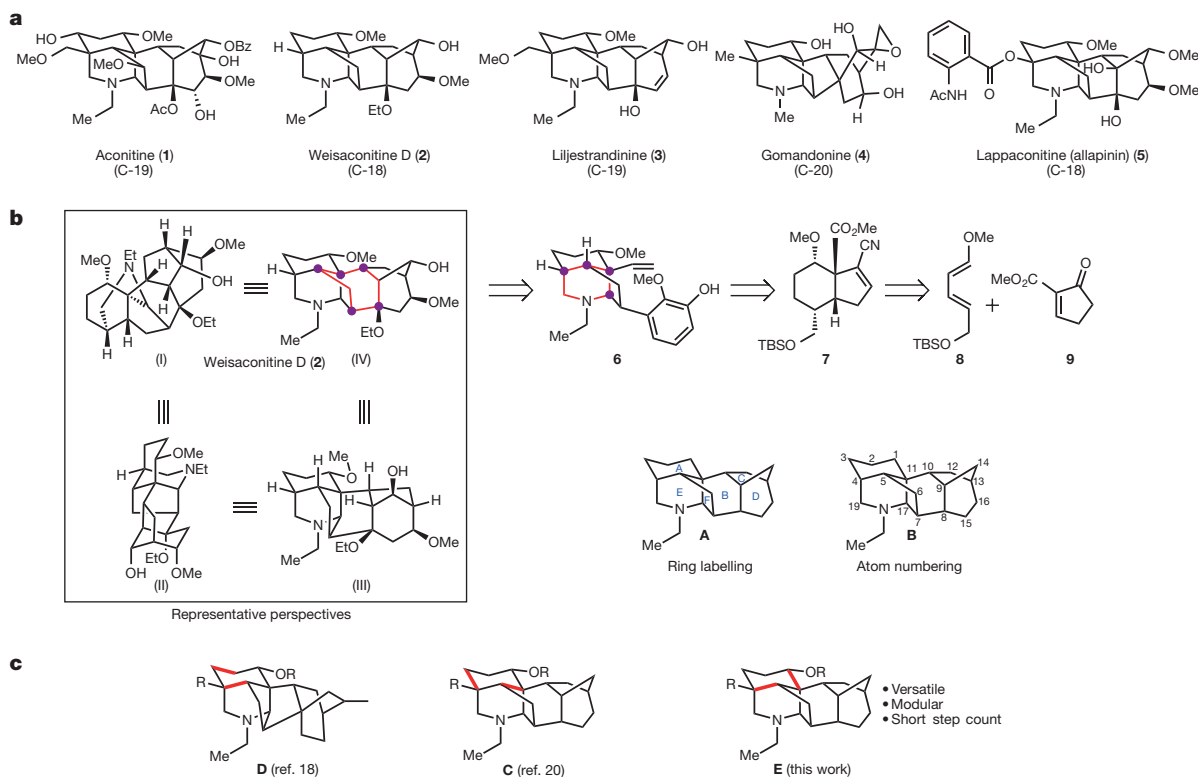


Figure 1 | Molecules referenced in this work and design strategy.

a, Selected C-18, C-19 and C-20 aconitine-type and denudatine-type diterpenoid alkaloids. **b**, Perspective drawings of weisaconitine D (left, boxed), retrosynthetic analysis highlighting maximally bridging

could be applied in the forward sense to forge the bicyclo[3.2.1] framework. In turn, identification of the piperidine ring in **6** as the maximally bridged ring for this compound triggered a retrosynthetic simplification by disconnection of the C19–N bond (see **B** in Fig. 1b for atom numbering) leading back to a bicycle that could be derived from **7**. Bicycle **7** was anticipated to be available from diene **8** and dienophile **9** using a Diels–Alder cycloaddition. Although alternative Diels–Alder cycloadditions (compare **C**, **D** and **E** in Fig. 1c) have been used in related total syntheses^{18–20}, the iterative application of network analysis, along with other retrosynthetic considerations such as the availability of starting materials and minimizing functional-group interconversions, led us to an alternative bond construction. Dehydro-hydrindane **7** possesses a variety of strategic synthetic handles that facilitate divergence in the synthetic scheme.

Similar retrosynthetic analyses can be proffered for the C-19 diterpenoid alkaloid liljestrandinine and for the C-20 alkaloid gomandonine (see Supplementary Information for more details). However, in these cases, the C4 bridgehead carbon would need to be quaternized, and **7** is suited for this purpose. From our analysis, **7** may also be used in the syntheses of other diterpenoid alkaloids of the hetidine, hetisine, denudatine and aconitine structural types (>900 members). Previously reported syntheses of diterpenoid alkaloids have mainly focused on specific targets or congeners in one family (for example, C-20 alkaloids)^{18–20,21}; our synthetic plan targets the range of C-18, C-19 and C-20 diterpenoid alkaloids.

Syntheses of weisaconitine D and liljestrandinine

The total synthesis of weisaconitine D was achieved in 30 steps from diene **8** and dienophile **9**, as outlined in the following. Our synthesis of weisaconitine D (Fig. 2a) commenced with the cycloaddition of known diene **8** (ref. 22) and cyclopentenone derivative **9** (ref. 23), yielding a cycloadduct that upon hydrogenation gives bicyclic ketone **10** (70%; 2 steps). Vinyl triflate formation and Pd(0)-catalysed cross-coupling with cyanide²⁴ yields α,β -unsaturated nitrile **7** (70%; 2 steps), which

served as a substrate for a Rh-catalysed conjugate addition with *in situ* generated lithium boronate **11**, to afford **12** in 60% yield. This conjugate addition step, which required careful optimization, provides a modular way to introduce the guaiacol derivative with high diastereoselectivity and enables access to various oxidation patterns on the C/D bicycle of the diterpenoid alkaloids by using other differently substituted arenes. Selective reduction of the ester group of **12** (in the presence of the cyano group) with Red-Al (ref. 25) and reoxidation of the resulting alcohol group to the aldehyde using the Dess–Martin periodinane reagent gives **13**. At this stage, Wittig olefination of the aldehyde group and hydration of the nitrile group using the conditions of ref. 26 provides carboxamide **14**. Hofmann rearrangement of the amide group and attendant trapping of the intermediate isocyanate with methanol, followed by fluoride-mediated cleavage of the *tert*-butyl dimethyl silyl (TBS) group gives **15**. Activation of the primary hydroxyl as the mesylate and exposure to KO^tBu effects alkylation to forge the C19–N bond and fashion the piperidine ring of **16** to complete the A, E and F rings (see **A**, Fig. 1b, for ring labelling) of the C-18 diterpenoid alkaloids. In preparation for the installation of the B, C and D rings, the methoxymethyl (MOM) group of **16** was removed and the resulting phenol subjected to oxidative dearomatization²⁷ to afford **17**. Dienone **17** smoothly undergoes intramolecular Diels–Alder cycloaddition upon heating to 150 °C to provide **18**, which is the core framework of the C-20 denudatine-type diterpenoid alkaloids (for example, gomandonine, **4**), bearing a bicyclo[2.2.2] moiety. The structure of this polycycle was secured by X-ray crystallographic analysis of benzoylated derivative **24** (Fig. 2b). In preparation for the transformation of the bicyclo[2.2.2] structural motif to the bicyclo[3.2.1] framework that is characteristic of the aconitine-type C-18 and C-19 alkaloids, the carbonyl group of **18** was reduced stereoselectively (presumably steered away from torsional strain with the β -disposed methoxy group of the dimethylketal), and the ketal was hydrolysed to unveil α -ketol **19**. Protection (MOM) of the secondary hydroxyl of **19** and diastereoselective reduction of

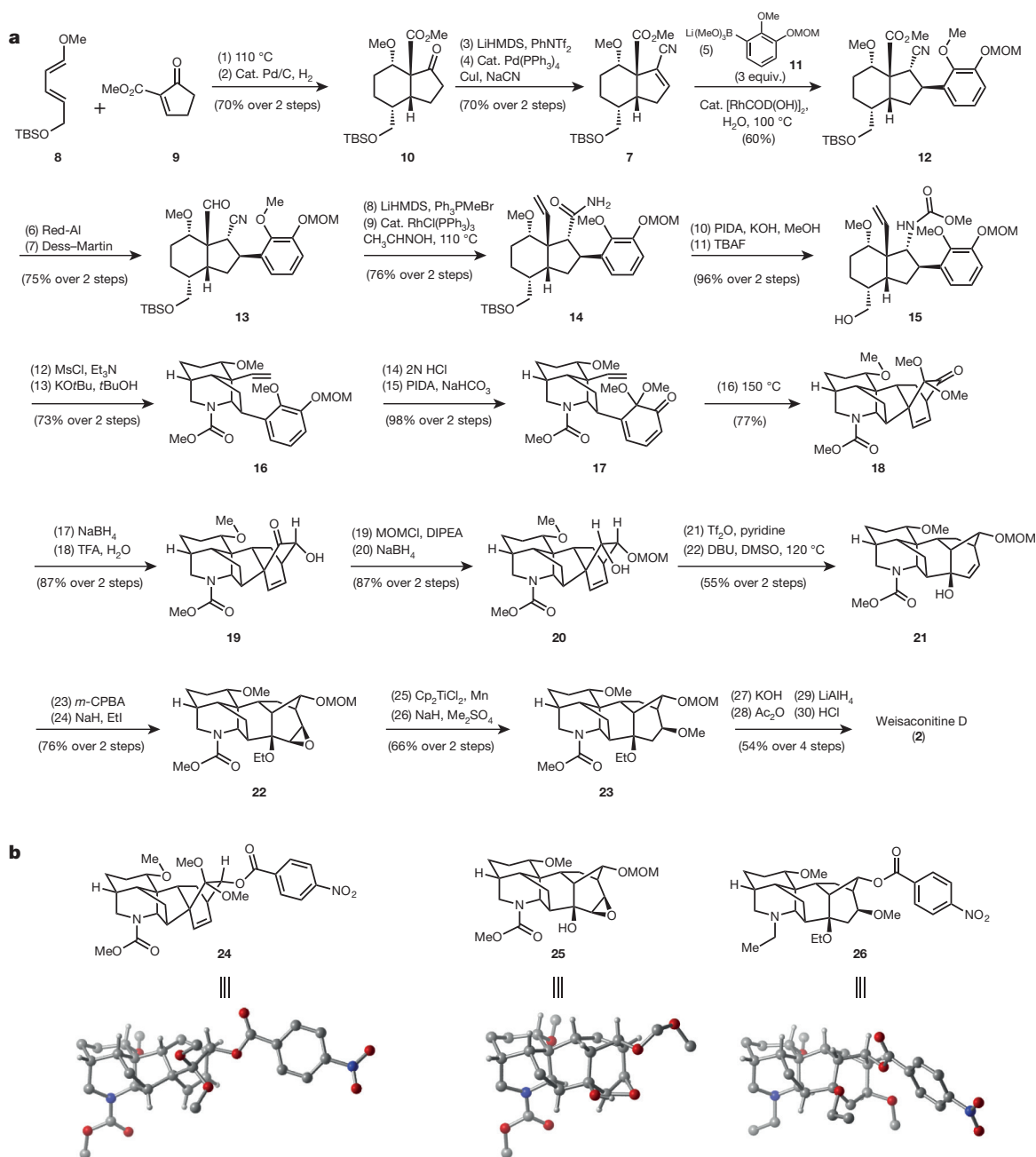


Figure 2 | Synthesis of weisaconitine D. **a**, Reaction sequence for the total synthesis of weisaconitine D. Reagents and conditions for each step are as follows: (1) **9** (1.0 equiv.), **8** (2.0 equiv.), toluene, 110 °C, 64 h; (2) Pd/C (10 wt%), H₂ gas (1 atm), EtOAc, room temperature (r.t.), 3 h, 70% yield over steps (1) and (2); (3) LiHMDS (lithium hexamethyldisilazide; 1.3 equiv.), PhNTf₂ (1.4 equiv.), THF, −78 °C to r.t., 12 h; (4) NaCN (2.2 equiv.), Pd(PPh₃)₄ (0.06 equiv.), CuI (0.12 equiv.), MeCN, reflux, 2 h, 70% yield over steps (3) and (4); (5) lithium boronate **11** (3.0 equiv.), [RhCOD(OH)]₂ (0.05 equiv.), dioxane/water, 16 h, 60% yield; (6) Red-Al (10 equiv.), CH₂Cl₂, −78 °C to r.t., 1 h, 82% yield; (7) Dess–Martin periodinane (2.0 equiv.), NaHCO₃ (5.0 equiv.), CH₂Cl₂, 0 °C, 1.5 h, 91% yield; (8) PPh₃MeBr (3.0 equiv.), LiHMDS (2.5 equiv.), THF, 0 °C to r.t., 1 h, 94% yield; (9) RhCl(PPh₃)₃ (0.3 equiv.), CH₃CHNOH/PhMe, reflux, 15 h, 81% yield; (10) KOH (3.4 equiv.), phenyliodonium diacetate (1.3 equiv.), MeOH, 0 °C to r.t., 3 h; (11) TBAF (tetrabutylammonium fluoride; 3.0 equiv.), THF, r.t., 5 h, 96% yield over steps (10) and (11); (12) MsCl (1.5 equiv.), CH₂Cl₂/Et₃N, 0 °C, 3 h; (13) KOtBu (3.0 equiv.), THF, 0 °C to r.t., 2 h, 73% yield over steps (12) and (13); (14) 2N HCl/*i*PrOH, 0 °C to r.t., 3.5 h, 99% yield; (15) phenyliodonium diacetate (1.5 equiv.),

NaHCO₃ (5.0 equiv.), MeOH, 0 °C, 1 h, 99% yield; (16) *p*-xylene, 150 °C, 17.5 h, 77% yield; (17) NaBH₄ (3.0 equiv.), MeOH, 0 °C to r.t., 3 h; (18) CHCl₃/TFA/water, 4 °C, 2 h, 87% yield over steps (17) and (18); (19) MOMCl (4.9 equiv.), DIPEA (N,N-diisopropylethylamine 10 equiv.), 4 °C to r.t., 16 h, 92% yield; (20) NaBH₄ (3.3 equiv.), MeOH, 4 °C, 2 h, 95% yield; (21) Tf₂O (10 equiv.), pyridine, CH₂Cl₂, −78 °C to r.t., 16 h; (22) DBU (3.3 equiv.), DMSO, 120 °C, 1 h, 55% yield over steps (21) and (22); (23) *m*-CPBA (5.2 equiv.), CH₂Cl₂, 0 °C to r.t., 16 h; (24) NaH (15 equiv.), EtI (15 equiv.), THF, 40 °C, 16 h, 76% yield over steps (23) and (24); (25) Cp₂TiCl₂ (2.2 equiv.), Mn (7.6 equiv.), H₂O (38 equiv.), THF, r.t., 16 h; (26) NaH (12 equiv.), Me₂SO₄ (7 equiv.), THF, 60 °C, 2 h, 66% yield over steps (25) and (26); (27) 4 M KOH, ethylene glycol, 100 °C, 120 h; (28) Ac₂O (9.4 equiv.), pyridine (28 equiv.), CH₂Cl₂, 0 °C to r.t., 16 h; and (29) LiAlH₄ (10 equiv.), Et₂O, 40 °C, 2 h; (30) 2N HCl, THF, 16 h, 54% yield over steps (27)–(30). Cat., catalyst. **b**, Images of intermediates **24** and **25**, and of derivatized weisaconitine D (**26**), created using CYLview⁴³. Most hydrogens (except stereocentres) have been removed for clarity. Hydrogen, white; carbon, grey; nitrogen, blue; oxygen, red.

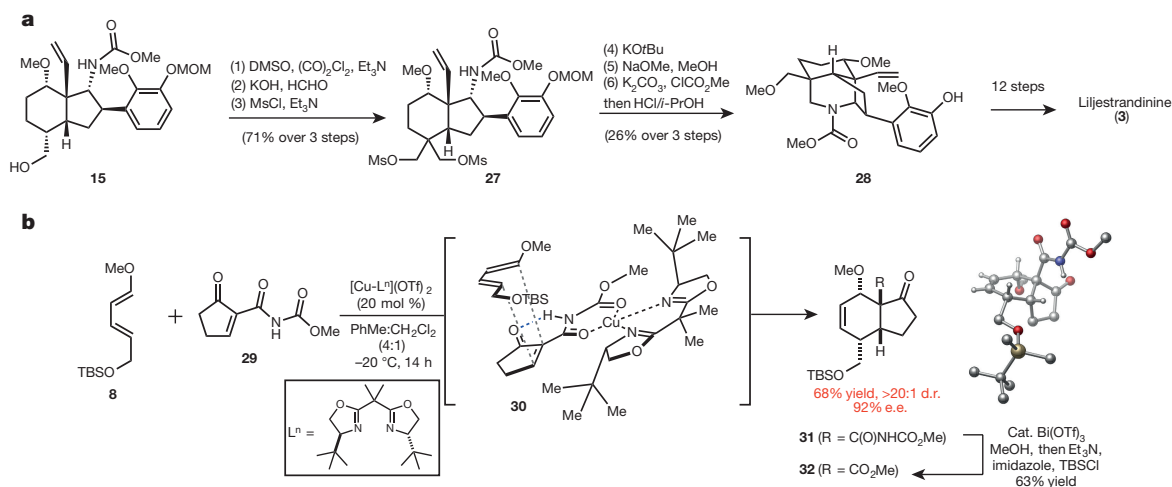


Figure 3 | Synthesis of liljestrandinine and enantioselective cycloaddition. **a**, Reaction sequence for the synthesis of liljestrandinine. Reagents and conditions for each step are as follows: (1) oxalyl chloride (2.9 equiv.), DMSO (6.2 equiv.), Et_3N (12 equiv.), CH_2Cl_2 , $-78^\circ C$ to r.t., 1 h, 95% yield; (2) formaldehyde (21 equiv.), 2N KOH, MeOH, r.t., 15 h, 96% yield; (3) $MsCl$ (3.5 equiv.), pyridine, $0^\circ C$ to r.t., 2 h, 78% yield; (4) $KOtBu$ (5 equiv.), THF, $50^\circ C$, 4 h; (5) 0.5 M NaOMe in MeOH, $120^\circ C$,

24 h; and (6) methyl chloroformate (20 equiv.), K_2CO_3 (40 equiv.), acetone, reflux, 20 h; 2N HCl, isopropanol, r.t., 4.5 h, 26% yield over steps (4)–(6). **b**, Enantioselective Diels–Alder cycloaddition approach using diene 8 and dienophile 29. Cat., catalyst. Image of 31 created using CYLview⁴³. Most hydrogens (except stereocentres) have been removed for clarity. Hydrogen, white; carbon, grey; nitrogen, blue; oxygen, red; silicon, yellow.

the ketone group provides alcohol 20. At this juncture, in preparation for a Wagner–Meerwein type rearrangement^{21,28}, the alcohol group of 20 was activated by triflation and, upon subsection of the triflate to 1,8-diazabicycloundec-7-ene (DBU) and DMSO, hexacycle 21 was isolated in 55% yield over the two steps. Although two isomeric allylic alcohols could result from the Wagner–Meerwein rearrangement, 21 is computed to be the more stable of the two (by $8.7 \text{ kcal mol}^{-1}$ (gas phase) and $8.4 \text{ kcal mol}^{-1}$ (DMSO) using density functional theory (M06-L/6-311G(*d,p*)) level of theory; see Supplementary Information for more details), presumably because it does not possess a strained bridgehead double bond. Several tactics were explored to achieve a formal hydro-methoxylation of the C15–C16 double bond, including the use of methanol in the presence of various protic and π -acids to activate the double bond²⁹, hydroboration (both inter- and intramolecular, directed by the secondary hydroxyl at C14 of 21 following MOM cleavage)³⁰ and variants of the hydration method presented in refs 31 and 32. Ultimately, the requisite methoxy group was installed at C16 of 21 using an epoxide intermediate. Thus, hydroxyl-directed epoxidation of the C15–C16 olefin group of 21 from the β -face using *meta*-chloroperbenzoic acid (*m*-CPBA) (see coloured model of 25, Fig. 2b) and ethylation of the tertiary hydroxyl yielded 22 (76% over 2 steps). Regioselective reductive opening of the epoxide using the conditions given in ref. 33 gave a β -disposed secondary alcohol group that was methylated to furnish 23 (66% over 2 steps). With the oxygenation of the D-ring of weisaconitine D secured, all that remained was to install the ethyl group on the piperidine nitrogen and to remove the MOM group to complete the synthesis. These tasks were accomplished in four steps: removal of the methoxycarbonyl (MOC) group of 23 (using KOH); acylation of the resultant secondary amine group (using Ac_2O); reduction of the acetamide (using $LiAlH_4$); and treatment with acid to remove the MOM group.

One key challenge that was not overcome in the previous syntheses of C-18, C-19 and C-20 diterpenoid alkaloids is how to achieve modular functionalization of the C4 position of the shared carbon framework. Here, we demonstrate that alcohol 15, a derivative of dehydro-hydrindane 7, can be used in the synthesis of the C-19 diterpenoid alkaloid liljestrandinine, which possesses a methoxymethylene group at C4 (Fig. 3a). Overall, the synthesis of liljestrandinine proceeds in 29 steps from diene 8 and dienophile 9, as summarized in the following. The primary hydroxyl of 15 was first oxidized, using Swern conditions, to the corresponding

aldehyde (not shown). Various attempts to alkylate the aldehyde enolate (as well as the enolates of related 6,5-bicycles) proved unfruitful and resulted in either non-specific decomposition or the addition of the electrophile from the undesired α -face (presumably due to developing syn-pentane interactions of the electrophile with the angular vinyl group). Ultimately, we found that an aldol–Cannizzaro sequence on the intermediate aldehyde, effected using KOH and formaldehyde, furnishes a geminal bis-methylene diol that was functionalized as the bis-mesylate (see 27 in Fig. 3a), where the C4 stereocentre is ablated. At this stage, alkylation of the carbamate nitrogen was accomplished (following ref. 34) with $KOtBu$ to forge the piperidine ring and reconstitute the C4 stereocentre (see 27 in Fig. 3a). Displacement of the remaining mesylate group with methoxide, reinstallation of the nitrogen protecting MOC group (which is partially cleaved during the methoxide displacement) and removal of the MOM group provides 28. Phenol 28 was advanced to an intermediate that is analogous to 21 (8 steps), and then to liljestrandinine using a sequence analogous to that described for 23 \rightarrow 2 (4 steps; see Supplementary Information for details).

An enantioselective Diels–Alder cycloaddition

The chemical syntheses of weisaconitine D and liljestrandinine described here rely on subsequent diastereoselective installation of all stereocentres from the four contiguous stereocentres that are introduced in the Diels–Alder reaction between diene 8 and dienophile 9. As such, a catalytic, enantioselective, Diels–Alder cycloaddition would enable enantioselective access to the natural products. In this regard, initial attempts to render the cycloaddition between 8 and 9 enantioselective with the aid of chiral, non-racemic, Lewis acid catalysts (for example, using the method of refs 35 and 36) resulted in low enantioselectivity and non-specific decomposition (primarily of diene 8 under the acidic conditions). Ultimately, 29 (ref. 37; for which we have developed a new, scalable synthesis; see Supplementary Information) was successfully used as a dienophile. This dienophile has enhanced reactivity because of an added intramolecular H bond³⁸ and a more highly organized transition state (see 30 in Fig. 3b for a model) that places the entio-discriminating substituents (for example, the *t*-butyl group of the bis-oxazoline ligand) proximal to the reacting dienophile double bond. A 68% yield of cycloadduct 31 (92% enantiomeric excess (e.e.); >20:1 diastereometric ratio (d.r.); see CYLview in Fig. 3b) was obtained using the conditions described

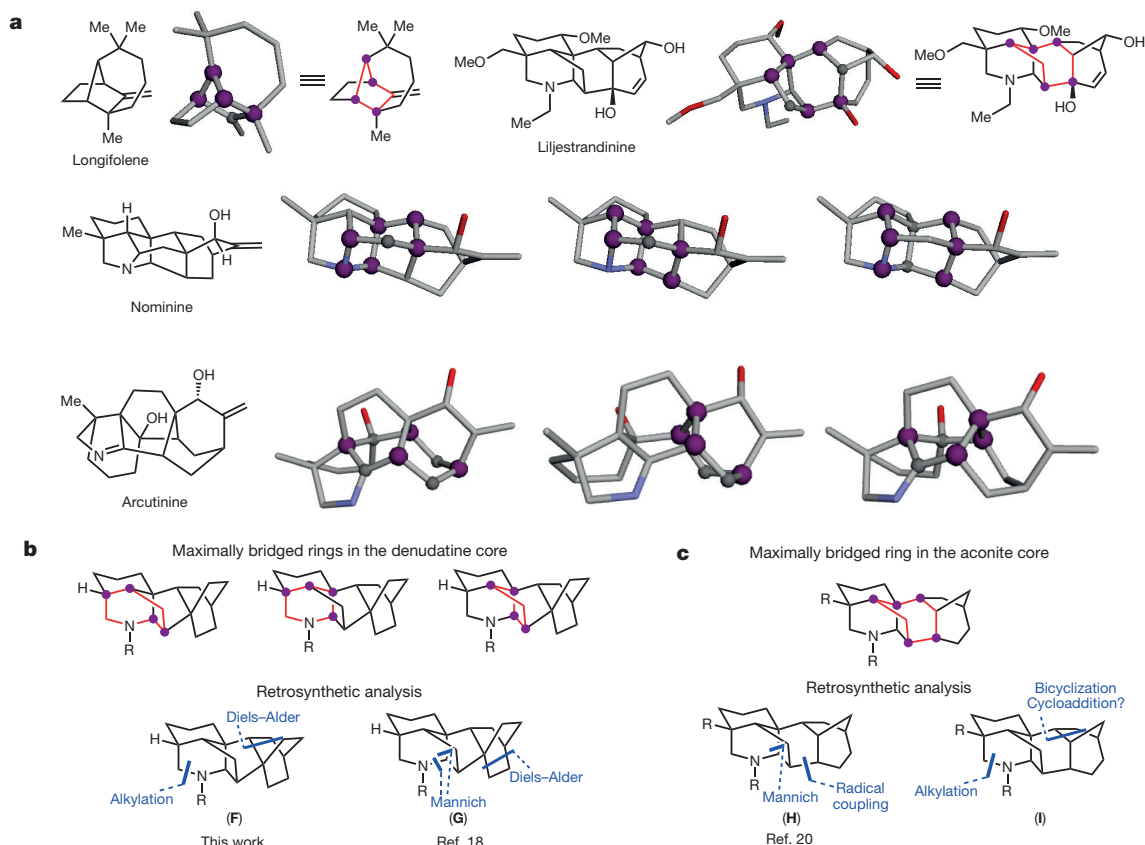


Figure 4 | Selected illustrations for network analysis graphing program. **a**, Selected molecules of a test set analysed using the newly developed graphing program to detect the maximally bridging ring. The program output is the '.pdb' image in grey. The maximally bridging ring is indicated by a combination of grey and purple spheres. The purple spheres represent bridgehead atoms in the maximally bridging ring and the grey spheres represent other atoms in the maximally bridging ring. ChemDraw renditions of the graphing program output are provided for

longifolene and liljestrandinine (top), nominine (middle) and arcutinine (bottom). For an extensive test set, see Supplementary Information; three-dimensional views of the output of the test set are located at <http://www.cadrerl.com/ring/>. **b**, ChemDraw renderings of the program output for the denudatine core and other key retrosynthetic disconnections applied here and in ref. 18. **c**, ChemDraw renderings of the program output for the aconite core and key retrosynthetic disconnections applied in ref. 20 and here.

in Fig. 3b. Furthermore, **31** is easily converted to **32**, which provides the enantio-enriched intermediate used in the racemic syntheses described in Figs 2a and 3a.

A web-based network analysis program

Our iterative application of network analysis¹³ to initiate a strategy for the syntheses of weisaconitine D and liljestrandinine led us to develop general ways to conduct such analyses. Previous implementations of network analysis in retrosynthesis, especially in the identification of the maximally bridged ring, have been carried out in a probabilistic manner, which invariably heightens the risk of errors^{39,40}. To overcome this shortcoming, we developed a web-based deterministic graphing program that permits the identification of the maximally bridged ring (or rings) for any molecule using the Chemistry Development Kit (CDK) software library^{41,42} (see Fig. 4a for the output of a test set; see Supplementary Information for more details). The algorithm we developed for this purpose is guaranteed to identify the maximally bridged ring (or rings) each time it is run. The program allows control of several criteria (for example, the number of atoms that comprise the maximally bridged ring or that span bridging atoms in the maximally bridged ring). It outputs the maximally bridged ring, or in the case of ties (for example, for nominine and arcutinidine in Fig. 4a), all maximally bridged rings.

Although many considerations are taken into account in retrosynthetic analyses of topologically complex molecules, network analysis often reveals strategic disconnections. For example, consider the denudatine core (Fig. 4b), which contains three rings that each possesses

four bridgehead atoms. By focusing on these rings for disconnection, maximum retrosynthetic simplification (that is, removal of bridging chains and fused rings) is achieved in the least number of steps with our approach (see **F** in Fig. 4b). A retrosynthetic analysis of the aconite framework, informed by network analysis (Fig. 4c) suggests that disconnections represented by **I** would provide maximum simplification. These latter strategic disconnections, which guided our approach to the syntheses of weisaconitine D and liljestrandinine, also indicate that a direct bicyclization to construct the bicyclo[3.2.1] moiety would provide the maximum benefit. Efforts to achieve this type of bicyclization are the subject of our ongoing studies. The creation of this web-based program should further facilitate the use of network analysis in developing retrosyntheses of other architecturally complex molecules and enable the identification of an efficient path to their syntheses.

Conclusion

The preparation of the denudatine core and total syntheses of weisaconitine D and liljestrandinine presented here reaffirm the utility of complex molecule synthesis as a driver for the implementation of chemical synthesis strategies that advance the field. Our approach offers a plan for the synthesis of a subset of C-18 and C-19 diterpenoid alkaloids and could enable access to related secondary metabolites including those in the C-20 family. The web-based deterministic graphing program we developed to analyse these topologically complex molecules, which builds on the work of ref. 13, should be useful in other contexts and might be valuable in the analysis and synthesis of other architecturally challenging molecules.

Received 25 August; accepted 30 October 2015.

Published online 16 December; corrected online 23 December 2015

(see full-text HTML for details).

- Nusim, S. H. (ed.) *Active Pharmaceutical Ingredients: Development, Manufacturing, and Regulation* 2nd edn, Vol. 205 of *Drugs and the Pharmaceutical Sciences* (CRC Press, 2009).
- Schaefer, B. *Natural Products in the Chemical Industry* 209–518 (Springer, 2014).
- Farina, V., Reeves, J. T., Senanayake, C. H. & Song, J. J. Asymmetric synthesis of active pharmaceutical ingredients. *Chem. Rev.* **106**, 2734–2793 (2006).
- dos Santos Pinheiro, E., Antunes, O. A. C. & Fortunak, J. M. D. A survey of the syntheses of active pharmaceutical ingredients for antiretroviral drug combinations critical to access in emerging nations. *Antiviral Res.* **79**, 143–165 (2008).
- Shenvi, R. A., O'Malley, D. P. & Baran, P. S. Chemoselectivity: the mother of invention in total synthesis. *Acc. Chem. Res.* **42**, 530–541 (2009).
- Hudlický, T. & Reed, J. W. *The Way of Synthesis* (Wiley-VCH, 2007).
- Service, R. F. Race for molecular summits. *Science* **285**, 184–187 (1999).
- Negishi, E.-i. Magical power of transition metals: past, present, and future (Nobel lecture). *Angew. Chem. Int. Ed.* **50**, 6738–6764 (2011).
- White, M. C. C–H bond functionalization and synthesis in the 21st century: a brief history and prospectus. *Synlett* **23**, 2746–2748 (2012).
- Kwok, R. Five hard truths for synthetic biology. *Nature* **463**, 288–290 (2010).
- MacMillan, J. & Beale, M. H. Diterpene biosynthesis. *Compr. Nat. Prod. Chem.* **2**, 217–243 (1999).
- Koehn, F. E. & Carter, G. T. The evolving role of natural products in drug discovery. *Nature Rev. Drug Discov.* **4**, 206–220 (2005).
- Corey, E. J., Howe, W. J., Orf, H. W., Pensak, D. A. & Petersson, G. General methods of synthesis analysis. Strategic bond disconnections for bridged polycyclic structures. *J. Am. Chem. Soc.* **97**, 6116–6124 (1975).
- Chan, T. Y. K. Aconite poisoning. *Clin. Toxicol.* **47**, 279–285 (2009).
- Catterall, W. A. et al. Voltage-gated ion channels and gating modifier toxins. *Toxicon* **49**, 124–141 (2007).
- Anger, T., Madge, D. J., Mulla, M. & Riddall, D. Medicinal chemistry of neuronal voltage-gated sodium channel blockers. *J. Med. Chem.* **44**, 115–137 (2001).
- Vakhitova, Yu. V. et al. A study of the mechanism of the antiarrhythmic action of allapinin. *Russ. J. Bioorg. Chem.* **39**, 92–101 (2013).
- Nishiyama, Y., Han-ya, Y., Yokoshima, S. & Fukuyama, T. Total synthesis of (–)-lepenine. *J. Am. Chem. Soc.* **136**, 6598–6601 (2014).
- Shi, Y., Wilms, J. T., Nordström, L. U., Tan, D. S. & Gin, D. Y. Total synthesis, relay synthesis, and structural confirmation of the C18-norditerpenoid alkaloid neofinaconitine. *J. Am. Chem. Soc.* **135**, 14313–14320 (2013).
- Wiesner, K., Tsai, T. Y. R., Huber, K., Bolton, S. E. & Vlahov, R. Total synthesis of talatisamine, a delphinine type alkaloid. *J. Am. Chem. Soc.* **96**, 4990–4992 (1974).
- Cherney, E. C., Lopchuk, J. M., Green, J. C. & Baran, P. S. A unified approach to ent-atrisane diterpenes and related alkaloids: synthesis of (–)-methyl atisenolate, (–)-isoatrisane, and the hetidine skeleton. *J. Am. Chem. Soc.* **136**, 12592–12595 (2014).
- Prabhakaran, J., Lhermitte, H., Das, J., Sasi-Kumar, T. K. & Grierson, D. S. The synthesis of a sulfone containing analogue of the esperamicin-A₁ aglycone: a hetero Diels–Alder approach. *Synlett* **5**, 658–662 (2000).
- Marx, J. N., Cox, J. H. & Norman, L. R. 2-Carbomethoxycyclopent-2-enone. *J. Org. Chem.* **37**, 4489–4491 (1972).
- Bandyopadhyaya, A. K. et al. Neurosteroid analogues. 15. A comparative study of the anesthetic and GABAergic actions of alphaxalone, Δ^{16} -alphaxalone and their corresponding 17-carbonitrile analogues. *Bioorg. Med. Chem. Lett.* **20**, 6680–6684 (2010).
- Yokota, S. & Miyamoto, S. Insecticide. Japanese patent 2008-024670 (2008).
- Lee, J., Kim, M., Chang, S. & Lee, H.-Y. Anhydrous hydration of nitriles to amides using aldioximes as the water source. *Org. Lett.* **11**, 5598–5601 (2009).
- Quideau, S. et al. Iodine-mediated and electrochemical oxidative transformations of 2-methoxy- and 2-methylphenols. *ARKIVOC* **2003**(vi), 106–119 (2003).
- Cheng, H., Xu, L., Chen, D.-L., Chen, Q.-H. & Wang, F.-P. Construction of the functionalized B/C/D ring system of C₁₉-diterpenoid alkaloids via intramolecular Diels–Alder reaction followed by Wagner–Meerwein rearrangement. *Tetrahedron* **68**, 1171–1176 (2012).
- Fürstner, A. & Davies, P. W. Catalytic carbophilic activation: catalysis by platinum and gold π acids. *Angew. Chem. Int. Ed.* **46**, 3410–3449 (2007).
- Evans, D. A., Fu, G. C. & Hoveyda, A. H. Rhodium(I)-catalyzed hydroboration of olefins. The documentation of regio- and stereochemical control in cyclic and acyclic systems. *J. Am. Chem. Soc.* **110**, 6917–6918 (1988).
- Isayama, S. & Mukaiyama, T. Novel method for the preparation of triethylsilyl peroxides from olefins by the reaction with molecular oxygen and triethylsilane catalyzed by bis(1,3-diketono)cobalt(II). *Chem. Lett.* **18**, 573–576 (1989).
- Isayama, S. An efficient method for the direct peroxygenation of various olefinic compounds with molecular oxygen and triethylsilane catalyzed by a cobalt(II) complex. *Bull. Chem. Soc. Jpn* **63**, 1305–1310 (1990).
- Cuerva, J. M. et al. Water: the ideal hydrogen-atom source in free-radical chemistry mediated by Ti^{III} and other single-electron transfer metals? *Angew. Chem. Int. Ed.* **45**, 5522–5526 (2006).
- Wiesner, K. Total synthesis of racemic talatisamine. *Pure Appl. Chem.* **41**, 93–112 (1975).
- Schötes, C. & Mezzetti, A. Asymmetric Diels–Alder reactions of unsaturated β -ketoesters catalyzed by chiral ruthenium PNNP complexes. *J. Am. Chem. Soc.* **132**, 3652–3653 (2010).
- Schötes, C., Althaus, M., Aardoom, R. & Mezzetti, A. Asymmetric Diels–Alder and Ficin reactions with alkylidene β -ketoesters catalyzed by chiral ruthenium PNNP complexes: mechanistic insight. *J. Am. Chem. Soc.* **134**, 1331–1343 (2012).
- Oyama, H., Orimoto, K., Niwa, T. & Nakada, M. Highly enantioselective catalytic asymmetric Mukaiyama–Michael reactions of cyclic α -alkylidene- β -oxo imides. *Tetrahedron Asym.* **26**, 262–270 (2015).
- Orimoto, K., Oyama, H., Namera, Y., Niwa, T. & Nakada, M. Catalytic asymmetric [4 + 2] cycloadditions and Hosomi–Sakurai reactions of α -alkylidene β -keto imides. *Org. Lett.* **15**, 768–771 (2013).
- Corey, E. J. & Cheng, X.-M. *The Logic of Chemical Synthesis* 43–44 (Wiley, 1989).
- Hoffmann, R. W. *Elements of Synthesis Planning* (Springer, 2009).
- Steinbeck, C. et al. The chemistry development kit (CDK): an open-source Java library for chemo- and bioinformatics. *J. Chem. Inform. Comput. Sci.* **43**, 493–500 (2003).
- Steinbeck, C. et al. Recent developments of the chemistry development kit (CDK) – an open-source Java library for chemo- and bioinformatics. *Curr. Pharm. Des.* **12**, 2111–2120 (2006).
- Legault, C. Y. CYLview, 1.0b. <http://www.cylview.org> (Université de Sherbrook, 2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements This project was supported by award no. RO1 GM084906 from the National Institute of General Medical Sciences. C.J.M. acknowledges a National Science Foundation graduate fellowship. G.M.G. is grateful to the NIH (5F31GM095238) for a graduate fellowship. T.P.L. and K.G.M.K. acknowledge postdoctoral fellowships from the NSERC (Canada). We are grateful to X.-Y. Liu for copies of ¹H and ¹³C NMR spectra for liljestrandinine (**3**). X-ray crystallography was performed by A. DiPasquale (NIH Shared Instrumentation Grant S10-RR027172). The AV-600, AV-500, DRX-500 and AVB-400 NMR instruments were partially supported by NIH grant SRR023679A, NIH grant 1S10RR016634-01, NSF grant CHE 9633007 and NSF grant CHE-0130862, respectively. We acknowledge E. Fisher (Pfizer) for contributions toward a practical synthesis of diene **8**. We thank K. Owens for help with generating '.sdf' files, and A. Chen, N. Kelly and K. Evens for the preparation of starting materials. We acknowledge M. Weber for computational calculations pertaining to allylic alcohol **21** and its isomer.

Author Contributions R.S. wrote the manuscript and all authors contributed to the reading and editing of the manuscript. C.J.M., G.M.G., J.C.L., T.P.L., S.K. and K.G.M.K. conducted the chemical reactions. S.K., K.G.M.K. and C.J.M. compiled the Supplementary Information. T.P.L. and G.M.G. made revisions and contributions to the Supplementary Information. J.Q., R.L. and R.S. conceptualized the graphing program, which was executed by J.Q. and R.L. J.Q., R.L. and R.S. prepared the portion of the Supplementary Information describing the graphing program. C.J.M. completed the synthesis and characterization of weisaconitine D, including the conversion of the [2.2.2] to the [3.2.1] bicycle (**19** → **21**), the formal hydromethoxylation sequence (**21** → **23**) and the development of robust conditions for the aryl conjugate addition (**7** → **12**; with K.G.M.K.) and Hofmann rearrangement (**13** → **14**). G.M.G. developed steps 3–17 in the synthesis of [2.2.2] bicycle **19**, including establishing a large-scale synthesis of **10**, synthesis of piperidine **16**, Diels–Alder cycloaddition of **17** and stereoselective reduction of ketone **18**. J.C.L. completed the synthesis of liljestrandinine, including conceptualization of the nitrile as a nitrogen atom surrogate, developing conditions for the conjugate addition of the functionalized arene (**7** → **12**) and establishing the sequence described for the conversion of **23** → **2** and **15** → **28**. T.P.L. contributed to the conceptualization of the synthetic route with substantial synthetic contributions made to the early portion of the synthesis including the establishment of a large-scale synthesis of **10**. S.K. developed the enantioselective Diels–Alder reaction (synthesis of **32**) and completed the optimization, scale-up and characterization of liljestrandinine. K.G.M.K. completed the optimization, scale-up and characterization of weisaconitine D, and optimized the conjugate addition (**7** → **12**; with C.J.M.) and the construction of piperidine **16**.

Author Information Crystallographic data can be obtained free of charge from the Cambridge Crystallographic Data Centre (CCDC; <https://summary.ccdc.cam.ac.uk/structure-summary-form>, reference numbers 1402704, 1402818, 1402820 and 1403763). The web-based graphing program we developed is available at <http://www.cadrl.com/maxbridge>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.S. (rsarpng@berkeley.edu).

Genome-wide patterns of selection in 230 ancient Eurasians

Iain Mathieson¹, Iosif Lazaridis^{1,2}, Nadin Rohland^{1,2}, Swapan Mallick^{1,2,3}, Nick Patterson², Songül Alpaslan Roodenberg⁴, Eadaoin Harney^{1,3}, Kristin Stewardson^{1,3}, Daniel Fernandes⁵, Mario Novak^{5,6}, Kendra Sirak^{5,7}, Cristina Gamba^{5,8,†}, Eppie R. Jones⁸, Bastien Llamas⁹, Stanislav Dryomov^{10,11}, Joseph Pickrell^{1,†}, Juan Luís Arsuaga^{12,13}, José María Bermúdez de Castro¹⁴, Eudald Carbonell^{15,16}, Fokke Gerritsen¹⁷, Aleksandr Khokhlov¹⁸, Pavel Kuznetsov¹⁸, Marina Lozano^{15,16}, Harald Meller¹⁹, Oleg Mochalov¹⁸, Vyacheslav Moiseyev²⁰, Manuel A. Rojo Guerra²¹, Jacob Roodenberg²², Josep Maria Vergès^{15,16}, Johannes Krause^{23,24}, Alan Cooper⁹, Kurt W. Alt^{19,25,26}, Dorcas Brown²⁷, David Anthony²⁷, Carles Lalueza-Fox²⁸, Wolfgang Haak^{9,23*}, Ron Pinhasi^{5*} & David Reich^{1,2,3*}

Ancient DNA makes it possible to observe natural selection directly by analysing samples from populations before, during and after adaptation events. Here we report a genome-wide scan for selection using ancient DNA, capitalizing on the largest ancient DNA data set yet assembled: 230 West Eurasians who lived between 6500 and 300 BC, including 163 with newly reported data. The new samples include, to our knowledge, the first genome-wide ancient DNA from Anatolian Neolithic farmers, whose genetic material we obtained by extracting from petrous bones, and who we show were members of the population that was the source of Europe's first farmers. We also report a transect of the steppe region in Samara between 5600 and 300 BC, which allows us to identify admixture into the steppe from at least two external sources. We detect selection at loci associated with diet, pigmentation and immunity, and two independent episodes of selection on height.

The arrival of farming in Europe around 8,500 years ago necessitated adaptation to new environments, pathogens, diets and social organizations. While indirect evidence of this adaptation can be detected in patterns of genetic variation in present-day people¹, these patterns are only echoes of past events, which are difficult to date and interpret, and are often confounded by neutral processes. Ancient DNA provides a direct way to study these patterns, and should be a transformative technology for studies of selection, just as it has transformed studies of human pre-history. Until now, however, the large sample sizes required to detect selection have meant that studies of ancient DNA have concentrated on characterizing effects at parts of the genome already believed to have been affected by selection^{2–5}.

Genome-wide ancient DNA from West Eurasia

We assembled genome-wide data from 230 ancient individuals from West Eurasia dated to between 6500 and 300 BC (Fig. 1a, Extended Data Table 1, Supplementary Data Table 1 and Supplementary Information section 1). To obtain this data set, we combined published data from 67 samples from relevant periods and cultures^{4–6}, with 163 samples for which we report new data, of which 83 have, to our knowledge,

never previously been analysed (the remaining 80 samples include 67 whose targeted single nucleotide polymorphism (SNP) coverage we tripled from 390,000 ('390k capture') to 1,240,000 ('1240k capture')⁷; and 13 with shotgun data for which we generated new data using our targeted enrichment strategy^{3,8}). The 163 samples for which we report new data are drawn from 270 distinct individuals who we screened for evidence of authentic DNA⁷. We used in-solution hybridization with synthesized oligonucleotide probes to enrich promising libraries for the targeted SNPs (Methods). The targeted sites include nearly all SNPs on the Affymetrix Human Origins and Illumina 610-Quad arrays, 49,711 SNPs on chromosome X, 32,681 SNPs on chromosome Y, and 47,384 SNPs with evidence of functional importance. We merged libraries from the same individual and filtered out samples with low coverage or evidence of contamination to obtain the final set of individuals. The 1240k capture gives access to genome-wide data from ancient samples with small fractions of human DNA and increases efficiency by targeting sites in the human genome that will actually be analysed. The effectiveness of the approach can be seen by comparing our results to the largest previously published ancient DNA study, which used a shotgun sequencing strategy⁵. Our median coverage on analysed SNPs

¹Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ³Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁴Independent researcher, Santpoort-Noord, The Netherlands. ⁵School of Archaeology and Earth Institute, Belfield, University College Dublin, Dublin 4, Ireland. ⁶Institute for Anthropological Research, Zagreb 10000, Croatia. ⁷Department of Anthropology, Emory University, Atlanta, Georgia 30322, USA. ⁸Smurfit Institute of Genetics, Trinity College Dublin, Dublin 2, Ireland. ⁹Australian Centre for Ancient DNA, School of Biological Sciences & Environment Institute, University of Adelaide, Adelaide, South Australia 5005, Australia. ¹⁰Laboratory of Human Molecular Genetics, Institute of Molecular and Cellular Biology, Siberian Branch of the Russian Academy of Sciences, Novosibirsk 630090, Russia. ¹¹Department of Paleolithic Archaeology, Institute of Archaeology and Ethnography, Siberian Branch of the Russian Academy of Sciences, Novosibirsk 630090, Russia. ¹²Centro Mixto UCM-ISCIII de Evolución y Comportamiento Humanos, 28040 Madrid, Spain. ¹³Departamento de Paleontología, Facultad Ciencias Geológicas, Universidad Complutense de Madrid, 28040 Madrid, Spain. ¹⁴Centro Nacional de Investigación sobre Evolución Humana (CENIEH), 09002 Burgos, Spain. ¹⁵IPHES. Institut Català de Paleoecologia Humana i Evolució Social, Campus Sescelades-URV, 43007 Tarragona, Spain. ¹⁶Area de Prehistoria, Universitat Rovira i Virgili (URV), 43002 Tarragona, Spain. ¹⁷Netherlands Institute in Turkey, Istiklal Caddesi, Nur-i Ziya Sokak 5, Beyoğlu 34433, Istanbul, Turkey. ¹⁸Volga State Academy of Social Sciences and Humanities, Samara 443099, Russia. ¹⁹State Office for Heritage Management and Archaeology Saxony-Anhalt and State Museum of Prehistory, D-06114 Halle, Germany. ²⁰Peter the Great Museum of Anthropology and Ethnography (Kunstkamera) RAS, St Petersburg 199034, Russia. ²¹Department of Prehistory and Archaeology, University of Valladolid, 47002 Valladolid, Spain. ²²The Netherlands Institute for the Near East, Leiden RA-2300, the Netherlands. ²³Max Planck Institute for the Science of Human History, D-07745 Jena, Germany. ²⁴Institute for Archaeological Sciences, University of Tübingen, D-72070 Tübingen, Germany. ²⁵Danube Private University, A-3500 Krems, Austria. ²⁶Institute for Prehistory and Archaeological Science, University of Basel, CH-4003 Basel, Switzerland. ²⁷Anthropology Department, Hartwick College, Oneonta, New York 13820, USA. ²⁸Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), 08003 Barcelona, Spain. [†]Present addresses: Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5–7, 1350 Copenhagen, Denmark (C.G.); New York Genome Center, New York, New York 10013, USA (J.P.).

*These authors contributed equally to this work.

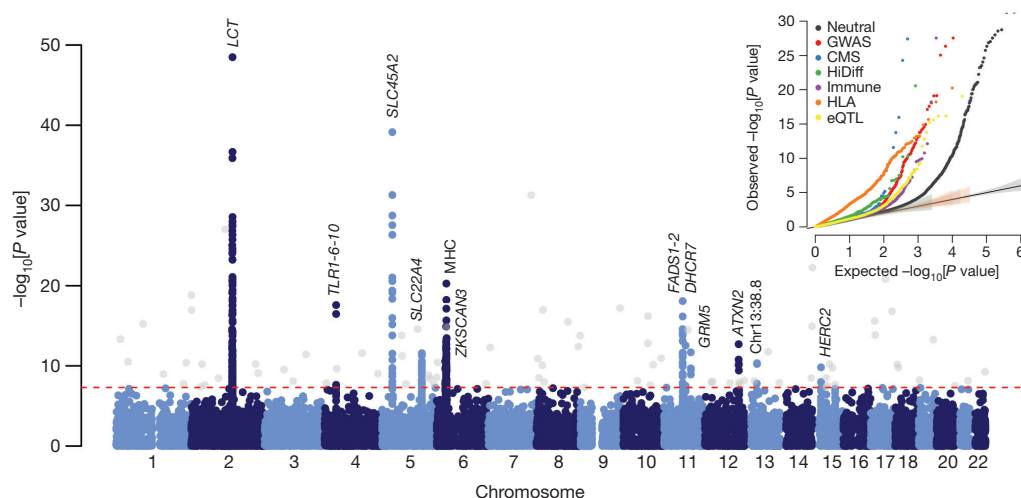


Figure 2 | Genome-wide scan for selection. GC-corrected $-\log_{10} P$ value for each marker (Methods). The red dashed line represents a genome-wide significance level of 0.5×10^{-8} . Genome-wide significant points filtered because there were fewer than two other genome-wide significant points within 1 Mb are shown in grey. Inset, quantile–quantile plots for corrected $-\log_{10} P$ values for different categories of potentially functional SNPs

(Methods). Truncated at $-\log_{10}[P \text{ value}] = 30$. All curves are significantly different from neutral expectation. CMS, composite of multiple signals selection hits; HiDiff, highly differentiated between HapMap populations; Immune, immune-related; HLA, human leukocyte antigen type tag SNPs; eQTL, expression quantitative trait loci (see Methods).

samples into three groups based on which of these populations they clustered with most closely (Fig. 1b and Extended Data Table 1). We estimated mixture proportions for the present-day European ancestry populations and tested every SNP to evaluate whether its present-day frequencies were consistent with this model. We corrected for test statistic inflation by applying a genomic control correction analogous to that used to correct for population structure in genome-wide association studies¹⁴. Of approximately one million non-monomorphic autosomal SNPs, the ~50,000 in the set of potentially functional SNPs were significantly more inconsistent with the model than neutral SNPs (Fig. 2), suggesting pervasive selection on polymorphisms of functional importance. Using a conservative significance threshold of $P = 5.0 \times 10^{-8}$, and a genomic control correction of 1.38, we identified 12 loci that contained at least three SNPs achieving genome-wide significance within 1 Mb of the most associated SNP (Fig. 2, Extended Data Table 3, Extended Data Fig. 3 and Supplementary Data Table 3).

The strongest signal of selection is at the SNP (rs4988235) responsible for lactase persistence in Europe^{15,16}. Our data (Fig. 3) strengthens previous reports that an appreciable frequency of lactase persistence in Europe only dates to the last 4,000 years^{3,5,17}. The allele's earliest appearance in the dataset is in a central European Bell Beaker sample (individual I0112) dated to between 2450 and 2140 BC. Two other independent signals related to diet are located on chromosome 11 near *FADS1* and *DHCR7*. *FADS1* and *FADS2* are involved in fatty acid metabolism, and variation at this locus is associated with plasma lipid and fatty acid concentration¹⁸. The selected allele of the most significant SNP (rs174546) is associated with decreased triglyceride levels¹⁸. This locus has experienced independent selection in non-European populations^{13,19,20} and is likely to be a critical component of adaptation to different diets. Variants at *DHCR7* and *NADSYN1* are associated with circulating vitamin D levels²¹ and the most associated SNP in our analysis, rs7940244, is highly differentiated across closely related northern European populations^{22,23}, suggesting selection related to variation in dietary or environmental sources of vitamin D.

Two signals have a potential link to coeliac disease. One occurs at the ergothioneine transporter *SLC22A4* that is hypothesized to have experienced a selective sweep to protect against ergothioneine deficiency in agricultural diets²⁴. Common variants at this locus are associated with increased risk for ulcerative colitis, coeliac disease, and irritable bowel

disease and may have hitchhiked to high frequency as a result of this sweep^{24–26}. However, the specific variant (rs1050152, L503F) that was thought to be the target did not reach high frequency until relatively recently (Extended Data Fig. 4). The signal at *ATXN2/SH2B3*—also associated with coeliac disease²⁵—shows a similar pattern (Extended Data Fig. 4).

The second strongest signal in our analysis is at the derived allele of rs16891982 in *SLC45A2*, which contributes to light skin pigmentation and is almost fixed in present-day Europeans but occurred at much lower frequency in ancient populations. In contrast, the derived allele of *SLC24A5* that is the other major determinant of light skin pigmentation in modern Europe (and that is not significant in the genome-wide scan for selection) appears fixed in the Anatolian Neolithic, suggesting that its rapid increase in frequency to around 0.9 in Early Neolithic Europe was mostly due to migration (Extended Data Fig. 4). Another pigmentation signal is at *GRM5*, where SNPs are associated with pigmentation possibly through a regulatory effect on nearby *TYR*²⁷. We also find evidence of selection for the derived allele of rs12913832 at *HERC2/OCA2*, which is at 100% frequency in the European hunter-gatherers we analysed, and is the primary determinant of light eye colour in present-day Europeans^{28,29}. In contrast to the other loci, the range of frequencies in modern populations is within that of ancient populations (Fig. 3). The frequency increases with higher latitude, suggesting a complex pattern of environmental selection.

The *TLR1–TLR6–TLR10* gene cluster is a known target of selection in Europe, possibly related to resistance to leprosy, tuberculosis or other mycobacteria^{30–32}. There is also a strong signal of selection at the major histocompatibility complex (MHC) on chromosome 6. The strongest signal is at rs2269424 near the genes *PPT2* and *EGFL8*, but there are at least six other apparently independent signals in the MHC (Extended Data Fig. 3); and the entire region is significantly more associated than the genome-wide average (residual inflation of 2.07 in the region on chromosome 6 between 29–34 Mb after genome-wide genomic control correction). This could be the result of multiple sweeps, balancing selection, or increased drift as a result of background selection reducing effective population size in this gene-rich region.

We find a surprising result in six Scandinavian hunter-gatherers (SHG) from Motala in Sweden. In three of six samples, we observe the haplotype carrying the derived allele of rs3827760 in the *EDAR*

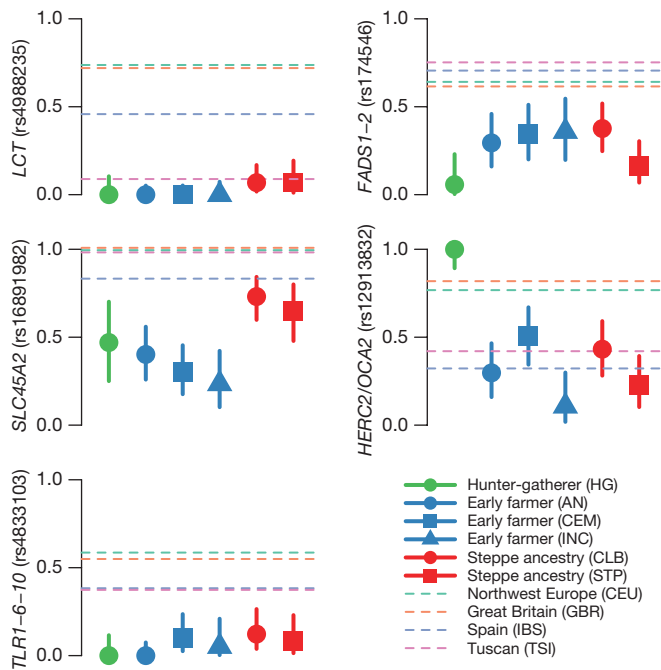


Figure 3 | Allele frequencies for five genome-wide significant signals of selection. Dots and solid lines show maximum likelihood frequency estimates and a 1.9-log-likelihood support interval for the derived allele frequency in each ancient population. Horizontal dashed lines show allele frequencies in the four modern 1000 Genomes populations. AN, Anatolian Neolithic; HG, hunter-gatherer; CEM, central European Early and Middle Neolithic; INC, Iberian Neolithic and Chalcolithic; CLB, central European Late Neolithic and Bronze Age; STP, steppe; CEU, Utah residents with northern and western European ancestry; IBS, Iberian population in Spain. The hunter-gatherer, early farmer and steppe ancestry classifications correspond approximately to the three populations used in the genome-wide scan with some differences (See Extended Data Table 1 for details).

gene (Extended Data Fig. 5), which affects tooth morphology and hair thickness^{33,34}, has been the target of a selective sweep in East Asia³⁵, and today is at high frequency in East Asians and Native Americans. The *EDAR* derived allele is largely absent in present-day Europe, except in Scandinavia, plausibly owing to Siberian movements into the region millennia after the date of the Motala samples. The SHG have no evidence of East Asian ancestry^{4,7}, suggesting that the *EDAR* derived allele may not have originated in the main ancestral population of East Asians as previously suggested³⁵. A second surprise is that, unlike closely related WHGs, the Motala samples

have predominantly derived pigmentation alleles at *SLC45A2* and *SLC24A5*.

Evidence of selection on height

We also tested for selection on complex traits. The best-documented example of this process in humans is height, for which the differences between northern and southern Europe have been driven by selection³⁶. To test for this signal in our data, we used a statistic that tests whether trait-affecting alleles are both highly correlated and more differentiated, compared to randomly sampled alleles³⁷. We predicted genetic heights for each population and applied the test to all populations together, as well as to pairs of populations (Fig. 4). Using 180 height-associated SNPs³⁸ (restricted to 169 for which we successfully obtained genotypes from at least two individuals from each population), we detect a significant signal of directional selection on height ($P = 0.002$). Applying this to pairs of populations allows us to detect two independent signals. First, the Iberian Neolithic and Chalcolithic samples show selection for reduced height relative to both the Anatolian Neolithic ($P = 0.042$) and the central European Early and Middle Neolithic ($P = 0.003$). Second, we detect a signal for increased height in the steppe populations ($P = 0.030$ relative to the central European Early and Middle Neolithic). These results suggest that the modern South–North gradient in height across Europe is due to both increased steppe ancestry in northern populations, and selection for decreased height in Early Neolithic migrants to southern Europe. We did not observe any other significant signals of polygenic selection in five other complex traits we tested: body mass index³⁹ ($P = 0.20$), waist-to-hip ratio⁴⁰ ($P = 0.51$), type 2 diabetes⁴¹ ($P = 0.37$), inflammatory bowel disease²⁶ ($P = 0.17$) and lipid levels¹⁸ ($P = 0.50$).

Future studies of selection with ancient DNA

Our results, which take advantage of the massive increase in sample size enabled by optimized techniques for sampling from the inner ear regions of the petrous bone, as well as in-solution enrichment methods for targeted SNPs, show how ancient DNA can be used to perform a genome-wide scan for selection. Our results also directly document selection on loci related to pigmentation, diet and immunity, painting a picture of populations adapting to settled agricultural life at high latitudes. For most of the signals, allele frequencies of modern Europeans are outside the range of any ancient populations, indicating that phenotypically, Europeans of 4,000 years ago were different in important respects from Europeans today, despite having overall similar ancestry. An important direction for future research is to increase the sample size for European selection scans (Extended Data Fig. 6), and to apply this approach to regions beyond Europe and to other species.

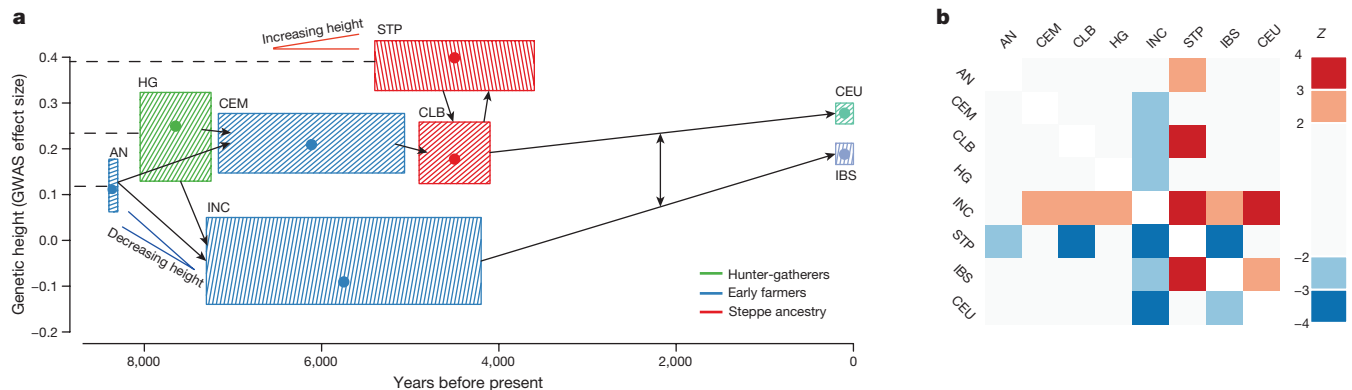


Figure 4 | Polygenic selection on height. **a**, Estimated genetic heights. Boxes show 0.05–0.95 posterior densities for population mean genetic height (Methods). Dots show the maximum likelihood point estimate. Arrows show major population relationships, dashed lines represent

ancestral populations. The symbols < and > label potentially independent selection events resulting in an increase or decrease in height. **b**, Z scores for the pairwise polygenic selection test. Positive if the column population is taller than the row population.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 March; accepted 30 October 2015.

Published online 23 November 2015.

1. Grossman, S. R. *et al.* Identifying recent adaptations in large-scale genomic data. *Cell* **152**, 703–713 (2013).
2. Wilde, S. *et al.* Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proc. Natl Acad. Sci. USA* **111**, 4832–4837 (2014).
3. Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory. *Nature Commun.* **5**, 5257 (2014).
4. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
5. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
6. Keller, A. *et al.* New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nature Commun.* **3**, 698 (2012).
7. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
8. Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225–228 (2014).
9. Pinhasi, R. *et al.* Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLoS ONE* **10**, e0129102 (2015).
10. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
11. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
12. Underhill, P. A. *et al.* The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *Eur. J. Hum. Genet.* **23**, 124–131 (2015).
13. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
14. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
15. Enattah, N. S. *et al.* Identification of a variant associated with adult-type hypolactasia. *Nature Genet.* **30**, 233–237 (2002).
16. Bersaglieri, T. *et al.* Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.* **74**, 1111–1120 (2004).
17. Burger, J., Kirchner, M., Bramanti, B., Haak, W. & Thomas, M. G. Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proc. Natl Acad. Sci. USA* **104**, 3736–3741 (2007).
18. Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
19. Fumagalli, M. *et al.* Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science* **349**, 1343–1347 (2015).
20. Mathias, R. A. *et al.* Adaptive evolution of the FADS gene cluster within Africa. *PLoS ONE* **7**, e44926 (2012).
21. Wang, T. J. *et al.* Common genetic determinants of vitamin D insufficiency: a genome-wide association study. *Lancet* **376**, 180–188 (2010).
22. Price, A. L. *et al.* The impact of divergence time on the nature of population structure: an example from Iceland. *PLoS Genet.* **5**, e1000505 (2009).
23. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
24. Huff, C. D. *et al.* Crohn's disease and genetic hitchhiking at IBD5. *Mol. Biol. Evol.* **29**, 101–111 (2012).
25. Hunt, K. A. *et al.* Newly identified genetic risk variants for celiac disease related to the immune response. *Nature Genet.* **40**, 395–402 (2008).
26. Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).
27. Beleza, S. *et al.* Genetic architecture of skin and eye color in an African-European admixed population. *PLoS Genet.* **9**, e1003372 (2013).
28. Sturm, R. A. *et al.* A single SNP in an evolutionary conserved region within intron 86 of the *HERC2* gene determines human blue-brown eye color. *Am. J. Hum. Genet.* **82**, 424–431 (2008).
29. Eiberg, H. *et al.* Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the *HERC2* gene inhibiting *OCA2* expression. *Hum. Genet.* **123**, 177–187 (2008).
30. Barreiro, L. B. *et al.* Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. *PLoS Genet.* **5**, e1000562 (2009).
31. Uciechowski, P. *et al.* Susceptibility to tuberculosis is associated with TLR1 polymorphisms resulting in a lack of TLR1 cell surface expression. *J. Leukoc. Biol.* **90**, 377–388 (2011).
32. Wong, S. H. *et al.* Leprosy and the adaptation of human toll-like receptor 1. *PLoS Pathog.* **6**, e1000979 (2010).
33. Fujimoto, A. *et al.* A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. *Hum. Mol. Genet.* **17**, 835–843 (2008).
34. Kimura, R. *et al.* A common variation in EDAR is a genetic determinant of shovel-shaped incisors. *Am. J. Hum. Genet.* **85**, 528–535 (2009).
35. Kamberov, Y. G. *et al.* Modeling recent human evolution in mice by expression of a selected EDAR variant. *Cell* **152**, 691–702 (2013).
36. Turchin, M. C. *et al.* Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nature Genet.* **44**, 1015–1019 (2012).
37. Berg, J. J. & Coop, G. *et al.* A population genetic signal of polygenic adaptation. *PLoS Genet.* **10**, e1004412 (2014).
38. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
39. Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genet.* **42**, 937–948 (2010).
40. Heid, I. M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nature Genet.* **42**, 949–960 (2010).
41. Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature Genet.* **44**, 981–990 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank P. de Bakker, J. Burger, C. Economou, E. Fornander, Q. Fu, F. Hallgren, K. Kirsanow, A. Mitnik, I. Olalde, A. Powell, P. Skoglund, S. Tabrizi and A. Tandon for discussions, suggestions about SNPs to include, or contribution to sample preparation or data curation. We thank S. Pääbo, M. Meyer, Q. Fu and B. Nickel for collaboration in developing the 1240k capture reagent. We thank J. M. V. Encinas and M. E. Prada for allowing us to resample La Braña 1. I.M. was supported by the Human Frontier Science Program LT001095/2014-L. C.G. was supported by the Irish Research Council for Humanities and Social Sciences (IRCHSS). F.G. was supported by a grant of the Netherlands Organization for Scientific Research, no. 380-62-005. A.K., P.K. and O.M. were supported by RFBR no. 15-06-01916 and RFH no. 15-11-63008 and O.M. by a state grant of the Ministry of Education and Science of the Russian Federation no. 33.1195.2014/k. J.K. was supported by ERC starting grant APGREID and DFG grant KR 4015/1-1. K.W.A. was supported by DFG grant AL 287 / 14-1. C.L.-F. was supported by a BFU2015-64699-P grant from the Spanish government. W.H. and B.L. were supported by Australian Research Council DP130102158. R.P. was supported by ERC starting grant ADNABIOARC (263441), and an Irish Research Council ERC support grant. D.R. was supported by US National Science Foundation HOMINID grant BCS-1032255, US National Institutes of Health grant GM100233, and the Howard Hughes Medical Institute.

Author Contributions W.H., R.P. and D.R. supervised the study. S.A.R., J.L.A., J.M.B., E.C., F.G., A.K., P.K., M.L., H.M., O.M., V.M., M.A.R., J.R., J.M.V., J.K., A.C., K.W.A., D.B., D.A., C.L., W.H., R.P. and D.R. assembled archaeological material. I.M., I.L., N.R., S.M., N.P., S.D., J.P., W.H. and D.R. analysed genetic data. N.R., E.H., K.St., D.F., M.N., K.Si., C.G., E.R.J., B.L., C.L. and W.H. performed wet laboratory ancient DNA work. I.M., I.L. and D.R. wrote the manuscript with input from all co-authors.

Author Information The aligned sequences are available through the European Nucleotide Archive under accession number PRJEB11450. The Human Origins genotype datasets including ancient individuals can be found at (http://genetics.med.harvard.edu/reich/Reich_Lab/Datasets.html). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to I.M. (iain.mathieson@hms.harvard.edu), W.H. (haak@shh.mpg.de), R.P. (ron.pinhasi@ucd.ie) or D.R. (reich@genetics.med.harvard.edu).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Ancient DNA analysis. We screened 433 next-generation sequencing libraries from 270 distinct samples for authentic ancient DNA using previously reported protocols⁷. All libraries that we included in nuclear genome analysis were treated with uracil-DNA-glycosylase (UDG) to reduce characteristic errors of ancient DNA⁴².

We performed in-solution enrichment for a targeted set of 1,237,207 SNPs using previously reported protocols^{4,7,43}. The targeted SNP set merges 394,577 SNPs first reported in ref. 7 (390k capture), and 842,630 SNPs first reported in ref. 44 (840k capture). For 67 samples for which we newly report data in this study, there was pre-existing 390k capture data⁷. For these samples, we only performed 840k capture and merged the resulting sequences with previously generated 390k data. For the remaining samples, we pooled the 390k and 840k reagents together to produce a single enrichment reagent, which we called 1240k. We attempted to sequence each enriched library up to the point where we estimated that it was economically inefficient to sequence further. Specifically, we iteratively sequenced more and more from each sample and only stopped when we estimated that the expected increase in the number of targeted SNPs hit at least once would be less than about one for every 100 new read pairs generated. After sequencing, we filtered out samples with <30,000 targeted SNPs covered at least once, with evidence of contamination based on mitochondrial DNA polymorphism⁴³, a high rate of heterozygosity on chromosome X despite being male⁴⁵, or an atypical ratio of X to Y sequences.

Of the targeted SNPs, 47,384 are 'potentially functional' sites chosen as follows (with some overlap): 1,290 SNPs identified as targets of selection in Europeans by the Composite of Multiple Signals (CMS) test¹; 21,723 SNPs identified as significant hits by genome-wide association studies, or with known phenotypic effect (GWAS); 1,289 SNPs with extremely differentiated frequencies between HapMap populations⁴⁶ (HiDiff); 9,116 'ImmunoChip' SNPs chosen for study of immunity-related phenotypes (Immune); 347 SNPs phenotypically relevant to South America (mostly altitude adaptation SNPs in *EGLN1* and *EPAS1*), 5,387 SNPs which tag HLA haplotypes and 13,672 expression quantitative trait loci⁴⁷ (eQTL). **Population history analysis.** We used two data sets for population history analysis. 'HO' consists of 592,169 SNPs, taking the intersection of the SNP targets and the Human Origins SNP array⁴; we used this data set for co-analysis of present-day and ancient samples. 'HOII' consists of 1,055,209 SNPs that additionally includes sites from the Illumina genotype array⁴⁸; we used this data set for analyses only involving the ancient samples.

On the HO data set, we carried out principal components analysis in smartpca⁴⁹ using a set of 777 West Eurasian individuals⁴, and projected the ancient individuals with the option 'lsqproject: YES'. We carried out admixture analysis on a set of 2,345 present-day individuals and the ancient samples after pruning for LD in PLINK 1.9 (<https://www.cog-genomics.org/plink2>)⁵⁰ with parameters '-indep-pairwise 200 25 0.4'. We varied the number of ancestral populations between $K = 2$ and $K = 20$, and used cross-validation (-cv.) to identify the value of $K = 17$ to plot in Extended Data Fig. 2f.

We used ADMIXTOOLS¹¹ to compute f -statistics, determining standard errors with a block jackknife and default parameters. We used the option 'inbreed: YES' when computing f_3 -statistics of the form $f_3(\text{ancient}; \text{Ref}_1, \text{Ref}_2)$ as the ancient samples are represented by randomly sampled alleles rather than by diploid genotypes. For the same reason, we estimated F_{ST} genetic distances between populations on the HO data set with at least two individuals in smartpca also using the 'inbreed: YES' option.

We estimated ancestral proportions as in Supplementary Information section 9 of ref. 7, using a method that fits mixture proportions on a 'test' population as a mixture of n 'reference' populations by using f_4 -statistics of the form $f_4(\text{test or ref}, O_1, O_2, O_3)$ that exploit allele frequency correlations of the test or reference populations with triples of outgroup populations. We used a set of 15 world outgroup populations^{4,7}. In Extended Data Fig. 2, we added WHG and EHG as outgroups for those analyses in which they are not used as reference populations. We plot $\text{resnorm} = \|\mathbf{t} - \mathbf{R}\hat{\mathbf{a}}\|_2^2$ the squared 2-norm of the residuals where $\hat{\mathbf{a}}$ is a vector of n estimated mixture proportions (summing to 1), \mathbf{t} is a vector of $m \binom{m-1}{2}$ f_4 -statistics of the form $f_4(\text{test}, O_1; O_2, O_3)$ for m outgroups, and \mathbf{R} is a $m \binom{m-1}{2} \times n$ matrix of the form $f_4(\text{ref}, O_1; O_2, O_3)$ (Supplementary Information section 9 of ref. 7).

We determined sex by examining the ratio of aligned reads to the sex chromosomes⁵¹. We assigned Y-chromosome haplogroups to males using version 9.1.129 of the nomenclature of the International Society of Genetic Genealogy (<http://www.isogg.org>), restricting analysis using samtools⁵² to sites with map quality and base quality of at least 30, and excluding two bases at the ends of each sequenced fragment.

Genome-wide scan for selection. For most ancient samples, we did not have sufficient coverage to make reliable diploid calls. We therefore used the counts of sequences covering each SNP to compute the likelihood of the allele frequency in each population. Suppose that at a particular site, for each population we have M samples with sequence level data, and N samples with full diploid genotype calls (Loschbour, Stuttgart and the 1,000 Genomes samples). For samples $i = 1 \dots N$, with diploid genotype data, we observe X copies of the reference allele out of $2N$ total chromosomes. For each of samples $i = (N+1) \dots (N+M)$, with sequence level data, we observe R_i sequences with the reference allele out of T_i total sequences. Then, the likelihood of the population reference allele frequency, p given data $D = \{X, N, R_i, T_i\}$ is given by

$$L(p; D) = B(X, 2N, p) \times \prod_{i=N+1}^{N+M} \left\{ p^2 B(R_i, T_i, 1-\varepsilon) + 2p(1-p) B(R_i, T_i, 0.5) + (1-p)^2 B(R_i, T_i, \varepsilon) \right\}$$

where $B(k, n, p) = \binom{n}{k} p^k (1-p)^{n-k}$ is the binomial probability distribution and

ε is a small probability of error, which we set to 0.001. We write $\ell(p; D)$ for the log-likelihood. To estimate allele frequencies, for example in Fig. 3 or for the polygenic selection test, we maximized this likelihood numerically for each population.

To scan for selection across the genome, we used the following test. Consider a single SNP. Assume that we can model the allele frequencies \mathbf{p}_{mod} in A modern populations as a linear combination of allele frequencies in B ancient populations \mathbf{p}_{anc} . That is, $\mathbf{p}_{mod} = \mathbf{C} \mathbf{p}_{anc}$, where \mathbf{C} is an A by B matrix with rows summing to 1. We have data D_j from population j which is some combination of sequence counts and genotypes as described above. Then, writing $\bar{\mathbf{p}} = [\mathbf{p}_{anc}, \mathbf{p}_{mod}] = [\mathbf{p}_1 \dots \mathbf{p}_{A+B}]$ the log-likelihood of the allele frequencies equals the sum of the log-likelihoods for each population.

$$\ell(\bar{\mathbf{p}}, \bar{D}) = \sum_{j=1}^{A+B} \ell(\mathbf{p}_j; D_j)$$

To detect deviations in allele frequency from expectation, we test the null hypothesis $H_0: \mathbf{p}_{mod} = \mathbf{C} \mathbf{p}_{anc}$ against the alternative $H_1: \mathbf{p}_{mod}$ unconstrained. We numerically maximize this likelihood in both the constrained and unconstrained model and use the fact that twice the difference in log-likelihood is approximately χ^2_A distributed to compute a test statistic and P value.

We defined the ancient source populations by the 'Selection group 1' label in Extended Data Table 1 and Supplementary Table 1 and used the 1000 Genomes CEU, GBR, IBS and TSI as the present-day populations. We removed SNPs that were monomorphic in all four of these modern populations as well as in 1000 Genomes Yoruba (YRI). We do not use FIN as one of the modern populations, because they do not fit this three-population model well. We estimated the proportions of (HG, EF, SA) to be CEU = (0.196, 0.257, 0.547), GBR = (0.362, 0.229, 0.409), IBS = (0, 0.686, 0.314) and TSI = (0, 0.645, 0.355). In practice, we found that there was substantial inflation in the test statistic, most likely due to unmodelled ancestry or additional drift. To address this, we applied a genomic control correction¹⁴, dividing all the test statistics by a constant, λ , chosen so that the median P value matched the median of the null χ^2_A distribution. Excluding sites in the potentially functional set, we estimated $\lambda = 1.38$ and used this value as a correction throughout. One limitation of this test is that, although it identifies likely signals of selection, it cannot provide much information about the strength or date of selection. If the ancestral populations in the model are, in fact, close to the real ancestral populations, then any selection must have occurred after the first admixture event (in this case, after 6500 BC), but if the ancestral populations are mis-specified, even this might not be true.

To estimate power, we randomly sampled allele counts from the full data set, restricting to polymorphic sites with a mean frequency across all populations of <0.1 . We then simulated what would happen if the allele had been under selection in all of the modern populations by simulating a Wright-Fisher trajectory with selection for 50, 100 or 200 generations, starting at the observed frequency. We took the final frequency from this simulation, sampled observations to replace the actual observations in that population, and counted the proportion of simulations that gave a genome-wide significant result after GC correction (Extended Data Fig. 6a). We resampled sequence counts for the observed distribution for each population to simulate the effect of increasing sample size, assuming that the coverage and distribution of the sequences remained the same (Extended Data Fig. 6b).

We investigated how the genomic control correction responded when we simulated small amounts of admixture from a highly diverged population (Yoruba;

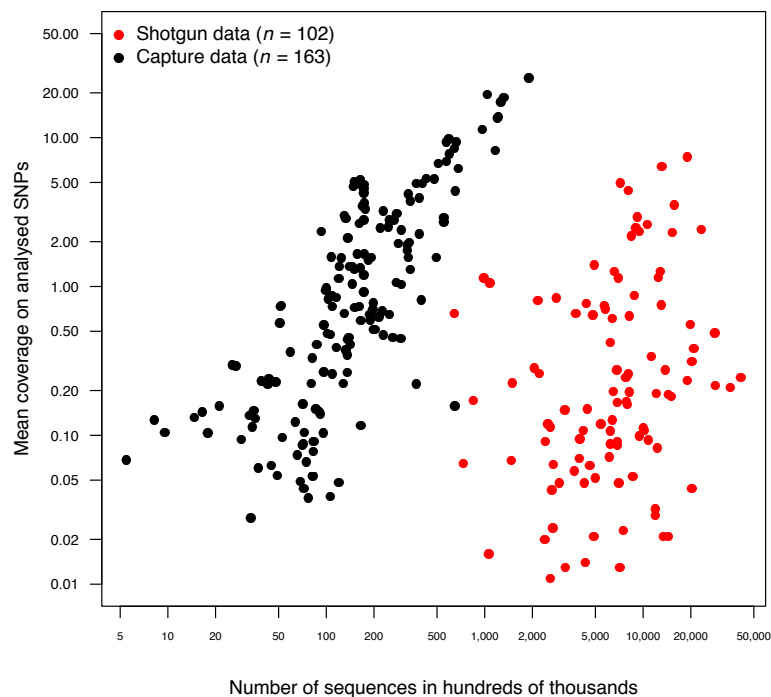
1000 Genomes YRI) into a randomly chosen modern population. The genomic inflation factor increases from around 1.38 to around 1.51 with 10% admixture, but there is little reduction in power (Extended Data Fig. 6c). Finally, we investigated how robust the test was to misspecification of the mixture matrix C . We re-ran the power simulations using a matrix $C' = xC + (1 - x)R$ for $x \in [0, 1]$ where R was a random matrix chosen so that for each modern population the mixture proportions of the three ancient populations were jointly uniformly distributed on $[0, 1]$. Increasing x increases the genomic inflation factor and reduces power, demonstrating the advantage of explicitly modelling the ancestries of the modern populations (Extended Data Fig. 6d).

Test for polygenic selection. We implemented the test for polygenic selection described by ref. 37. This evaluates whether trait-associated alleles, weighted by their effect size, are over-dispersed compared to randomly sampled alleles, in the directions associated with the effects measured by genome-wide association studies (GWAS). For each trait, we obtained a list of significant SNP associations and effect estimates from GWAS data, and then applied the test both to all populations combined and to selected pairs of populations. For height, we restricted the list of GWAS associations to 169 SNPs where we observed at least two chromosomes in all tested populations (Selection population 2). We estimated frequencies in each population by computing the maximum likelihood estimate (MLE), using the likelihood described above. For each test we sampled SNPs, frequency-matched in 20 bins, computed the test statistic Q_X and for ease of comparison converted these to Z scores, signed according to the direction of the genetic effects. Theoretically Q_X has a χ^2 distribution but in practice, it is over-dispersed. Therefore, we report bootstrap P values computed by sampling 10,000 sets of frequency-matched SNPs.

To estimate population-level genetic height in Fig. 4a, we assumed a uniform prior on $[0, 1]$ for the frequency of all height-associated alleles, and then sampled from the posterior joint frequency distribution of the alleles, assuming they were independent, using a Metropolis–Hastings sampler with a $N(0, 0.001)$ proposal density. We then multiplied the sampled allele frequencies by the effect sizes to get a distribution of genetic height.

Code availability. Code implementing the selection analysis is available at https://github.com/mathii/europe_selection.

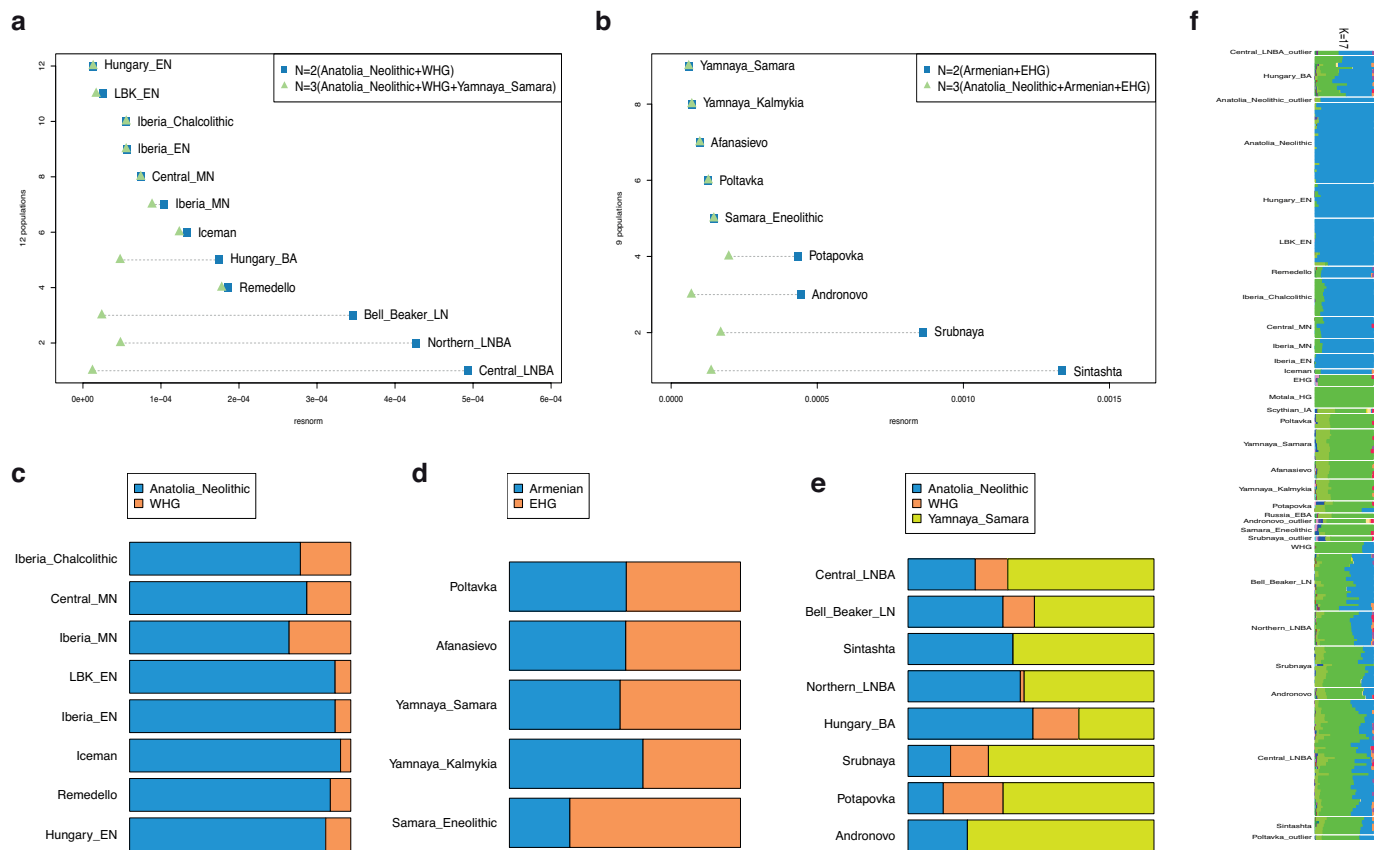
42. Briggs, A. W. *et al.* Removal of deaminated cytosines and detection of *in vivo* methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010).
43. Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110**, 2223–2227 (2013).
44. Fu, Q. *et al.* An early modern human from Romania with a recent Neanderthal ancestor. *Nature* **524**, 216–219 (2015).
45. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).
46. International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
47. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511 (2013).
48. Li, J. Z. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104 (2008).
49. Loh, P. R. *et al.* Inferring admixture histories of human populations using linkage disequilibrium. *Genetics* **193**, 1233–1254 (2013).
50. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4** (2015).
51. Skoglund, P., Storå, J., Götherström, A. & Jakobsson, M. Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* **40**, 4477–4482 (2013).
52. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
53. Norton, H. L. *et al.* Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. *Mol. Biol. Evol.* **24**, 710–722 (2007).
54. Bokor, S. *et al.* Single nucleotide polymorphisms in the *FADS* gene cluster are associated with delta-5 and delta-6 desaturase activities estimated by serum fatty acid ratios. *J. Lipid Res.* **51**, 2325–2333 (2010).
55. Tanaka, T. *et al.* Genome-wide association study of plasma polyunsaturated fatty acids in the InCHIANTI Study. *PLoS Genet.* **5**, e1000338 (2009).
56. Ahn, J. *et al.* Genome-wide association study of circulating vitamin D levels. *Hum. Mol. Genet.* **19**, 2739–2745 (2010).
57. Gründemann, D. *et al.* Discovery of the ergothioneine transporter. *Proc. Natl Acad. Sci. USA* **102**, 5256–5261 (2005).
58. Chauhan, S. *et al.* ZKSCAN3 is a master transcriptional repressor of autophagy. *Mol. Cell* **50**, 16–28 (2013).
59. Soler Artigas, M. *et al.* Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nature Genet.* **43**, 1082–1090 (2011).
60. Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).



	Shotgun data (Literature)	Capture data (Newly reported)
Average number of raw read pairs (hundreds of thousands, similar to cost in dollars)	8730	241
Median coverage on analyzed SNPs	0.19	0.75

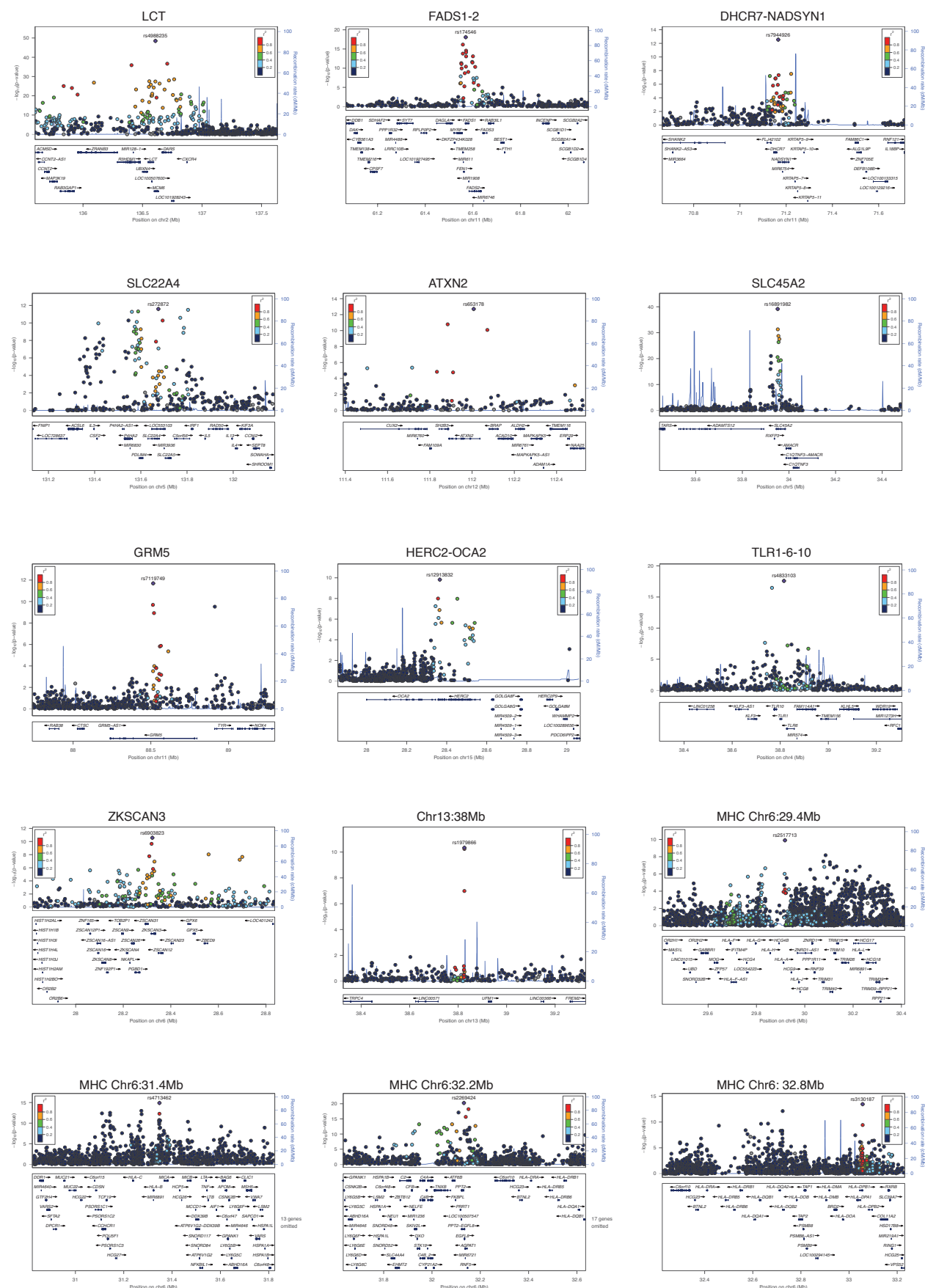
Extended Data Figure 1 | Efficiency and cost-effectiveness of 1240k capture. We plot the number of raw sequences against the mean coverage of analysed SNPs after removal of duplicates, comparing the 163 samples

for which capture data are newly reported in this study, against the 102 samples analysed by shotgun sequencing in ref. 5. We caution that the true cost is more than that of sequencing alone.



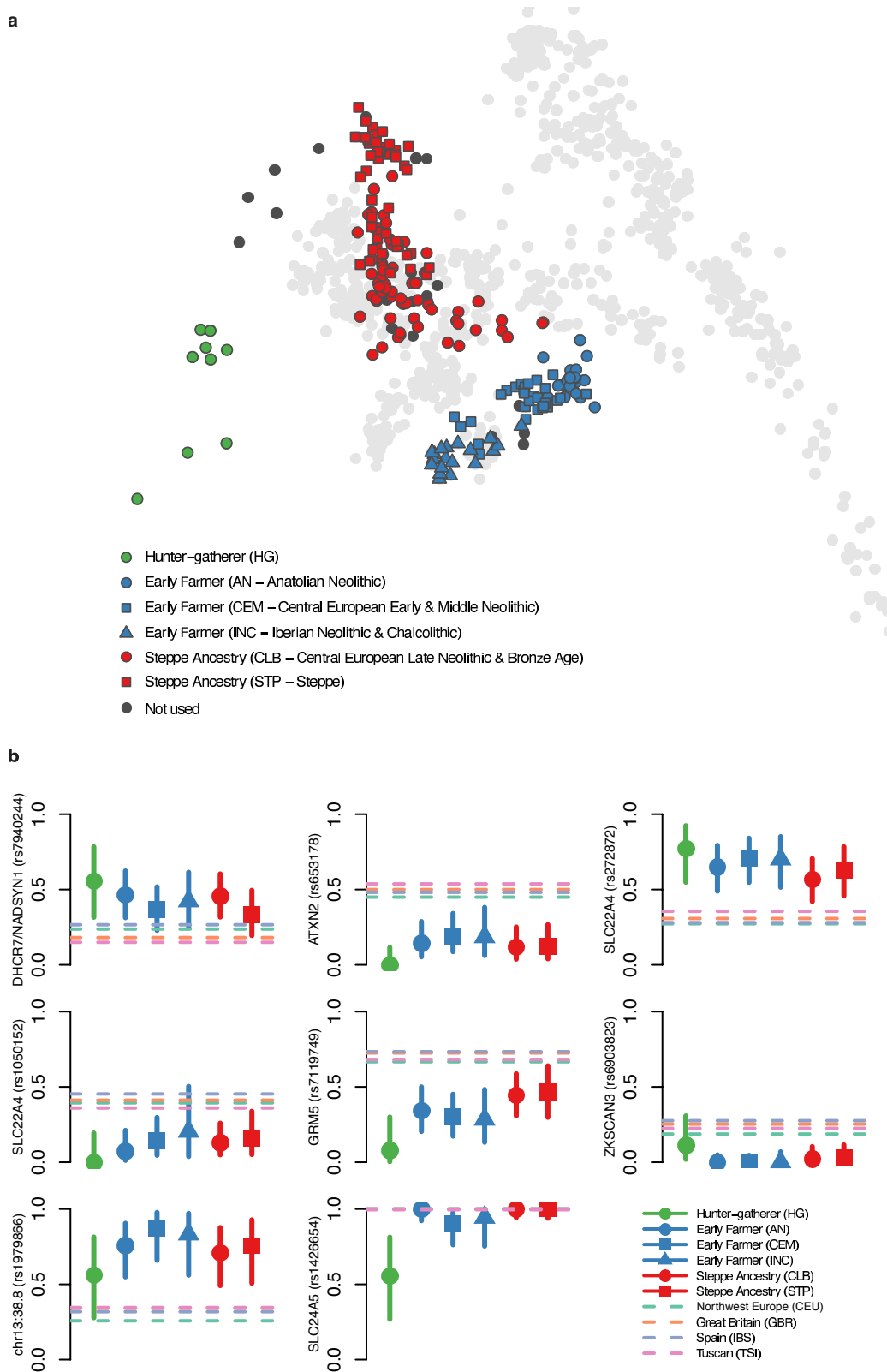
Extended Data Figure 2 | Early isolation and later admixture between farmers and steppe populations. **a**, Mainland European populations later than 3000 BC are better modelled with steppe ancestry as a third ancestral population, (closer correspondence between empirical and estimated f_4 -statistics as estimated by resnorm; Methods). **b**, Later (post-Poltavka) steppe populations are better modelled with Anatolian Neolithic as a third ancestral population. **c**, Estimated mixture proportions of mainland

European populations without steppe ancestry. **d**, Estimated mixture proportions of Eurasian steppe populations without Anatolian Neolithic ancestry. **e**, Estimated mixture proportions of later populations with both steppe and Anatolian Neolithic ancestry. **f**, Admixture plot at $k = 17$ showing population differences over time and space. EN, Early Neolithic; MN, Middle Neolithic; LN, Late Neolithic; BA, Bronze Age; LNBA, Late Neolithic and Bronze Age.



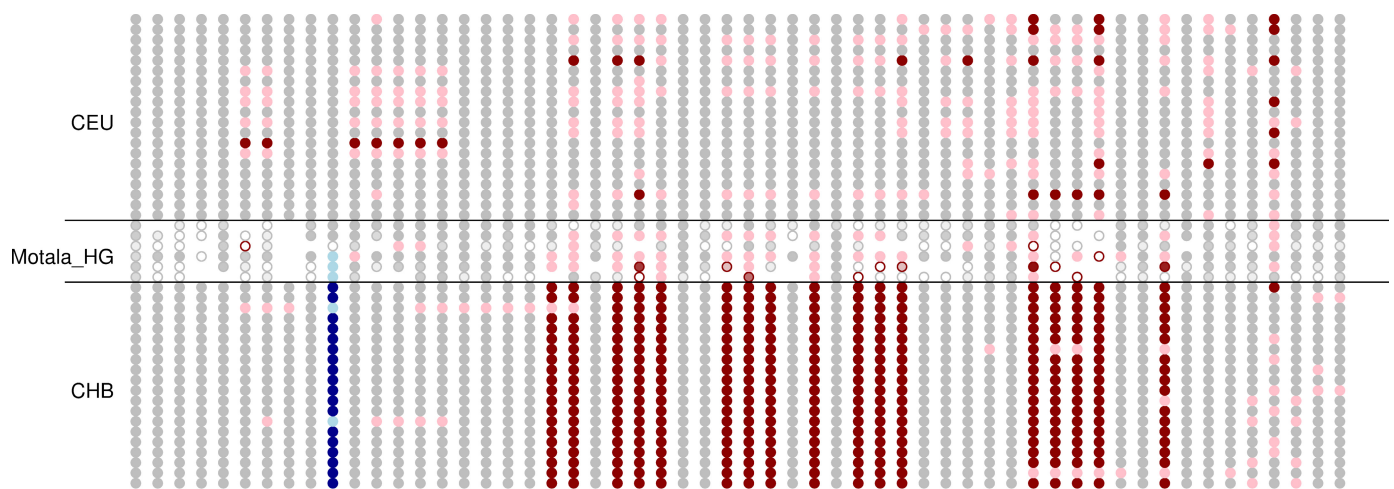
Extended Data Figure 3 | Regional association plots. LocusZoom⁶⁰ plots for genome-wide significant signals. Points show the $-\log_{10} P$ value for each SNP, coloured according to their linkage disequilibrium

(LD; units of r^2) with the most associated SNP. The blue line shows the recombination rate, with scale on right hand axis in centimorgans per megabase (cM/Mb). Genes are shown in the lower panel of each subplot.



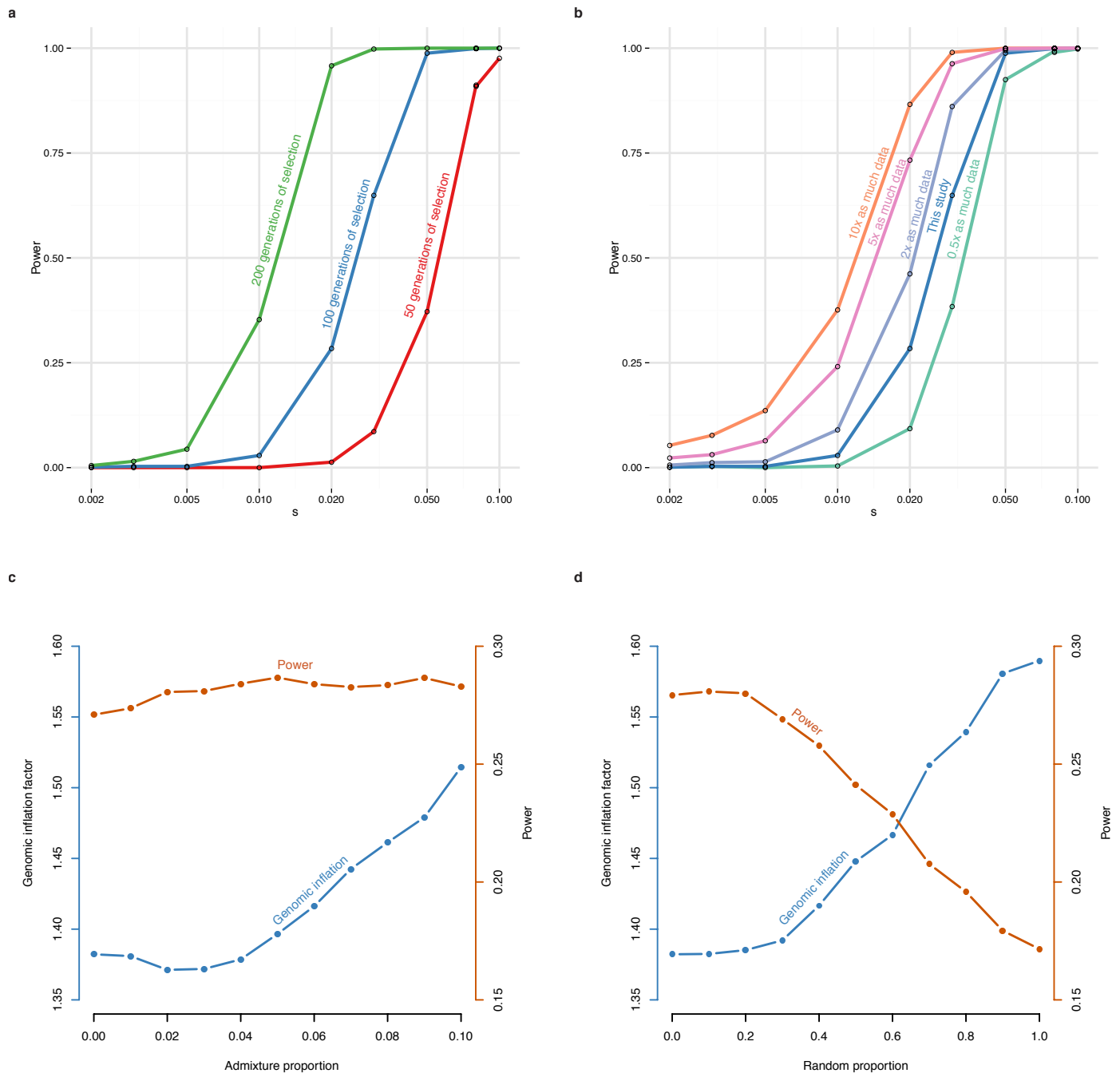
Extended Data Figure 4 | PCA of selection populations and derived allele frequencies for genome-wide significant signals. a, Ancient samples projected onto principal components of modern samples, as in Fig. 1, but labelled according to selection populations defined in Extended Data Table 1. **b,** Allele frequency plots as in Fig. 3. Six signals

not included in Fig. 3—for *SLC22A4* we show both rs272872, which is our strongest signal, and rs1050152, which was previously hypothesized to be under selection, and we also show *SLC24A5*, which is not genome-wide significant but is discussed in the main text.



Extended Data Figure 5 | Motala haplotypes carrying the derived, selected *EDAR* allele. This figure compares the genotypes at all sites within 150 kb of rs3827760 (in blue) for the 6 Motala samples and 20 randomly chosen CHB (Chinese from Beijing) and CEU (Utah residents with northern and western European ancestry) samples. Each row is a sample and each column is a SNP. Grey means homozygous for the major (in CEU) allele. Pink denotes heterozygous and red indicates homozygous for the other allele. For the Motala samples, an open circle means that

there is only a single sequence, otherwise the circle is coloured according to the number of sequences observed. Three of the Motala samples are heterozygous for rs3827760 and the derived allele lies on the same haplotype background as in present-day East Asians. The only other ancient samples with evidence of the derived *EDAR* allele in this data set are two Afanasievo samples dating to 3300–3000 BC, and one Scythian dating to 400–200 BC (not shown).



Extended Data Figure 6 | Estimated power of the selection scan.

a, Estimated power for different selection coefficients (s) for a SNP that is selected in all populations for either 50, 100 or 200 generations. **b**, Effect of increasing sample size, showing estimated power for a SNP selected for 100 generations, with different amounts of data, relative to the main text. **c**, Effect of admixture from Yoruba (YRI) into one of the

modern populations, showing the effect on the genomic inflation factor (blue, left axis) and the power to detect selection on a SNP selected for 100 generations with a selection coefficient of 0.02. **d**, Effect of mis-specification of the mixture proportions. Here 0 on the x axis corresponds to the proportions we used, and 1 corresponds to a random mixture matrix.

Extended Data Table 1 | 230 ancient individuals analysed in this study

By population	Population	Date range	N	Out	Rel	Eff N Chr	Selection population 1	Selection population 2
	WHG	8.2-8.0 kya	3	0	0	4.66	HG	HG
	Motala_HG	7.9-7.5 kya	6	0	0	5.19	HG	HG
	Anatolia_Neolithic	8.4-8.3 kya	24	1	1	22.49	EF	AN
	Hungary_EN	7.7-7.7 kya	10	0	0	8.81	EF	CEM
	LBK_EN	7.5-7.1 kya	15	0	0	11.15	EF	CEM
	Central_MN	5.9-5.8 kya	6	0	0	3.66	EF	CEM
	Iberia_EN	7.3-7.2 kya	4	0	1	3.54	EF	INC
	Iberia_MN	5.9-5.6 kya	4	0	0	3.47	EF	INC
	Iberia_Chalcolithic	4.8-4.2 kya	12	0	2	5.93	EF	INC
	Remedello	5.5-5.1 kya	3	0	0	0.93	EF	-
	Iceman	5.4-5.1 kya	1	0	0	1.90	EF	-
	Central_LNBA	4.9-4.6 kya	35	1	2	17.55	SA	CLB
	Yamnaya_Samara	5.4-4.9 kya	9	0	0	6.55	SA	STP
	Yamnaya_Kalmykia	5.3-4.7 kya	6	0	0	3.50	SA	STP
	Afanasievo	5.3-5.0 kya	5	0	0	3.01	SA	STP
	Poltavka	4.9-4.7 kya	4	1*	0	4.28	SA	STP
	Sintashta	4.3-4.1 kya	5	0	0	2.35	SA	STP
	Potapovka	4.2-4.1 kya	3	0	0	0.66	SA	STP
	Srubnaya	3.9-3.6 kya	12	1*	1	7.68	SA	STP
	Andronovo	3.8-3.6 kya	3	1*	0	3.87	SA	STP
	Russia_EBA	4.9-4.5 kya	1	0	0	0.21	SA	-
	Northern_LNBA	4.9-4.5 kya	10	0	0	3.81	SA	-
	Bell_Beaker_LN	4.5-4.5 kya	17	0	1	6.64	SA^	CLB
	Hungary_BA	4.2-4.1 kya	12	0	0	4.18	SA^	CLB
	EHG	7.7-7.6 kya	3	0	0	2.15	-	-
	Samara_Eneolithic	7.2-6.0 kya	3	0	0	1.07	-	-
	Scythian_IA	2.4-2.2 kya	1	0	0	1.26	-	-

By selection population	Selection population 1	Date range	N	Out	Rel	Eff N Chr	Description
	EF	8.4-4.2 kya	79	0	0	61.88	Early Farmer
	HG	8.2-7.5 kya	9	0	0	9.85	Hunter-gatherer
	SA	5.4-3.6 kya	93	3	0	52.14	Steppe Ancestry
	Selection population 2	Date range	N	Out	Rel	Eff N Chr	Description
	AN	8.4-8.3 kya	24	0	0	22.49	Anatolian Neolithic
	CEM	7.7-5.8 kya	31	0	0	23.62	Central European Early and Middle Neolithic
	INC	7.3-4.2 kya	20	0	0	12.95	Iberian Neolithic and Chalcolithic
	HG	8.2-7.5 kya	9	0	0	9.85	Hunter-gatherer
	CLB	4.9-4.1 kya	64	0	0	28.38	Central European Late Neolithic and Bronze Age
	STP	5.4-3.6 kya	47	3	0	30.58	Steppe

Population, samples grouped by a combination of date, source, archaeology and genetics; Date range, approximate date range of samples in this group; N, number of individuals sampled; Out, number of PCA outliers (marked with an asterisk if used in selection analysis); Rel, number of related individuals removed; Eff N Chr, average over sites of the effective number of chromosomes when we use genotype likelihoods, computed as 2 per called site for samples with genotype calls, or $2 - 0.5^{c-1}$ for samples with read depth c ; Selection population 1, coarse population labels (marked with a caret if not used in genome-wide scan); Selection population 2, fine population labels. E/M/LN, Early/Middle/Late Neolithic; LBK, *Linearbandkeramik*; E/S/WHG, Eastern/Scandinavian/Western hunter-gatherer; EBA, Early Bronze Age; IA, Iron Age.

Extended Data Table 2 | Key f_4 -statistics used to support claims about population history

A	B	C	D	$f_4(\text{A, B, C, D})$	Z	Number of SNPs	Interpretation
Anatolia_Neolithic	LBK_EN	WHG	Chimp	-0.00114	-6.8	1003751	Early European Farmers had more WHG ancestry than Anatolian Neolithic
Anatolia_Neolithic	Hungary_EN	WHG	Chimp	-0.00212	-11.9	929553	
Anatolia_Neolithic	Iberia_EN	WHG	Chimp	-0.00244	-9.6	904437	
Iberia_EN	Iberia_Chalcolithic	WHG	Chimp	-0.00311	-10.5	802471	Iberian Chalcolithic had more WHG ancestry than Iberian Early Neolithic
Iberia_MN	Iberia_Chalcolithic	WHG	Chimp	0.00010	0.3	779905	Iberian Chalcolithic did not have more WHG ancestry than Iberian Middle Neolithic
EHG	Samara_Eneolithic	MA1	Chimp	0.00140	2.3	463388	First dilution of Ancient North Eurasian ancestry (prior to the Bronze Age Yamnaya culture)
EHG	Yamnaya_Samara	MA1	Chimp	0.00513	10.6	645211	
Samara_Eneolithic	Yamnaya_Samara	MA1	Chimp	0.00366	7.6	482492	
EHG	Yamnaya_Samara	Armenian	Chimp	-0.00191	-6.1	547370	Contribution of Near Eastern ancestry to the Bronze Age Yamnaya culture
EHG	Yamnaya_Kalmykia	Armenian	Chimp	-0.00180	-5.4	536989	
Samara_Eneolithic	Yamnaya_Samara	Armenian	Chimp	-0.00100	-3.3	405599	
EHG	Poltavka	Armenian	Chimp	-0.00175	-4.9	541983	
Yamnaya_Samara	Yamnaya_Kalmykia	MA1	Chimp	-0.00010	-0.3	675630	Stability of Ancient North Eurasian ancestry between Early Bronze Age Yamnaya from Kalmykia and Samara, and the Middle Bronze Age Poltavka
Yamnaya_Samara	Poltavka	MA1	Chimp	-0.00014	-0.4	673726	
Yamnaya_Kalmykia	Poltavka	MA1	Chimp	0.00012	0.3	659346	
Yamnaya_Samara	Srubnaya	MA1	Chimp	0.00151	5.1	691149	Second dilution of Ancient North Eurasian ancestry (prior to the Late Bronze Age Srubnaya culture)
Yamnaya_Kalmykia	Srubnaya	MA1	Chimp	0.00161	4.8	676735	
Poltavka	Srubnaya	MA1	Chimp	0.00164	4.5	674756	
Yamnaya_Samara	Srubnaya	LBK_EN	Chimp	-0.00225	-11.4	974659	Arrival of Early European Farmer-related ancestry prior to the Late Bronze Age Srubnaya culture. Statistics with Anatolia_Neolithic instead of LBK_EN are similar ($Z < -8$, not shown).
Yamnaya_Kalmykia	Srubnaya	LBK_EN	Chimp	-0.00264	-11.4	951827	
Poltavka	Srubnaya	LBK_EN	Chimp	-0.00210	-9.0	948968	
EHG	Yamnaya_Samara	Armenian	LBK_EN	-0.00080	-5.0	559478	Different source of dilution of Ancient North Eurasian ancestry prior to the Yamnaya (Near Eastern) vs. prior to the Srubnaya (Early European Farmer-related)
EHG	Yamnaya_Kalmykia	Armenian	LBK_EN	-0.00086	-5.2	548882	
EHG	Poltavka	Armenian	LBK_EN	-0.00069	-4.1	553996	
Yamnaya_Samara	Srubnaya	Armenian	LBK_EN	0.00138	13.1	585240	
Yamnaya_Kalmykia	Srubnaya	Armenian	LBK_EN	0.00142	11.3	574333	
Poltavka	Srubnaya	Armenian	LBK_EN	0.00134	10.7	577082	
Ref ₁	Ref ₂	Test	$f_3(\text{Test}; \text{Ref1, Ref2})$	Z	Number of SNPs	Interpretation	
WHG	Anatolia_Neolithic	Hungary_EN	-0.00412	-6.7	548445	Early European farmers were formed by admixture between Anatolia Neolithic and WHG (the non-significant signal in the Iberia_EN may be due to genetic drift specific to this population)	
WHG	Anatolia_Neolithic	LBK_EN	-0.00257	-4.6	654357		
WHG	Anatolia_Neolithic	Iberia_EN	0.00179	1.4	389101		
EHG	Armenian	Poltavka	-0.00539	-3.9	213055	Early and Middle Bronze Age steppe pastoralists were formed by admixture between EHG and a population of Near Eastern ancestry	
EHG	Armenian	Yamnaya_Kalmykia	-0.00537	-4.2	213996		
EHG	Armenian	Yamnaya_Samara	-0.00586	-6.2	276568		
LBK_EN	Yamnaya_Samara	Srubnaya	-0.00630	-11.2	584111	Srubnaya was formed by admixture between populations related to Yamnaya and Early European Farmers	

Extended Data Table 3 | Twelve genome-wide significant signals of selection

Lead SNP	Chromosome	Position (hg19)	P Value	Range (Mb)	Genes	Potential function
rs4988235	2	136,608,646	3.19E-49	135.3-137.3	<i>MCM6,LCT</i>	Lactase persistence ^{*15}
rs16891982	5	33,951,693	7.05E-40	33.8-34.0	<i>SLC45A2</i>	Skin pigmentation ^{*53}
rs2269424	6	32,132,233	5.41E-21	29.9-33.1	MHC region	Immunity [*]
rs174546	11	61,569,830	8.18E-19	61.5-61.6	<i>FADS1,FADS2</i>	Fatty acid metabolism ^{*18,54,55}
rs4833103	4	38,815,502	2.58E-18	38.7-38.8	<i>TLR1,TLR6,TLR10</i>	Immunity ^{31,32}
rs653178	12	112,007,756	1.96E-13	111.9-112.6	<i>ATXN2,SH2B3</i>	Unknown
rs7944926	11	71,165,625	2.86E-13	71.2-71.2	<i>DHCR7,NADSYN1</i>	Vitamin D metabolism ^{21,56}
rs7119749	11	88,515,022	2.03E-12	88.5-88.9	<i>GRM5</i>	Skin pigmentation ²⁷
rs272872	5	131,675,864	2.56E-12	131.4-131.8	<i>SLC22A4</i>	Ergothioneine transport ⁵⁷
rs6903823	6	28,322,296	2.96E-11	28.3-28.7	<i>ZKSCAN3,ZSCAN31</i>	Autophagy ⁵⁸ , Lung function ⁵⁹
rs1979866	13	38,825,900	4.60E-11	38.1-38.8	-	Unknown
rs12913832	11	28,365,618	1.5E-10	29.9-30.1	<i>HERC2,OCA2</i>	Eye color ^{*28,29}

Chromosome/Position/Range, co-ordinates (hg19) of the SNP with the most significant signal, and the approximate range in which genome-wide significant SNPs are found. Genes, genes in which the top SNP is located, and selected nearby genes. Potential function, function of the gene, or specific trait under selection. Marked with an asterisk if the signal was still genome-wide significant in an analysis that used only the populations that correspond best to the three ancestral populations (WHG, Anatolian Neolithic and Bronze Age steppe), resulting in a less powerful test with the effective number of chromosomes analysed at the average SNP reduced from 125 to 50, a genomic control correction of 1.32, and five genome-wide significant loci that are a subset of the original twelve. Refs 53–59 are cited in this table.

Complete nitrification by *Nitrospira* bacteria

Holger Daims¹, Elena V. Lebedeva², Petra Pjevac¹, Ping Han¹, Craig Herbold¹, Mads Albertsen³, Nico Jehmlich⁴, Marton Palatinszky¹, Julia Vierheilig¹, Alexandr Bulaev², Rasmus H. Kirkegaard³, Martin von Bergen^{4,5}, Thomas Rattei⁶, Bernd Bendinger⁷, Per H. Nielsen³ & Michael Wagner¹

Nitrification, the oxidation of ammonia via nitrite to nitrate, has always been considered to be a two-step process catalysed by chemolithoautotrophic microorganisms oxidizing either ammonia or nitrite. No known nitrifier carries out both steps, although complete nitrification should be energetically advantageous. This functional separation has puzzled microbiologists for a century. Here we report on the discovery and cultivation of a completely nitrifying bacterium from the genus *Nitrospira*, a globally distributed group of nitrite oxidizers. The genome of this chemolithoautotrophic organism encodes the pathways both for ammonia and nitrite oxidation, which are concomitantly expressed during growth by ammonia oxidation to nitrate. Genes affiliated with the phylogenetically distinct ammonia monooxygenase and hydroxylamine dehydrogenase genes of *Nitrospira* are present in many environments and were retrieved on *Nitrospira*-contigs in new metagenomes from engineered systems. These findings fundamentally change our picture of nitrification and point to completely nitrifying *Nitrospira* as key components of nitrogen-cycling microbial communities.

Nitrification is catalysed by ammonia-oxidizing bacteria (AOB)¹ or archaea (AOA)² and nitrite-oxidizing bacteria (NOB)¹. Since the pioneering studies by Winogradsky more than a century ago³, nitrifying microorganisms are generally perceived as specialized chemolithoautotrophs that obtain energy for growth by oxidizing either ammonia or nitrite. The known ammonia-oxidizing microbes (AOM) and NOB are phylogenetically not closely related, and none of these organisms can oxidize both substrates. This separation of the two nitrification steps in different organisms leads to a tight cross-feeding interaction and the frequently observed co-aggregation of AOM with NOB in nitrifying consortia⁴. However, the functional separation is a puzzling phenomenon since complete nitrification would yield more energy ($\Delta G^\circ = -349 \text{ kJ mol}^{-1} \text{ NH}_3$) than either single step ($\Delta G^\circ = -275 \text{ kJ mol}^{-1} \text{ NH}_3$ for ammonia oxidation to nitrite and $\Delta G^\circ = -74 \text{ kJ mol}^{-1} \text{ NO}_2^-$ for nitrite oxidation to nitrate). Thus, an organism catalysing complete nitrification should have growth advantages over the ‘incomplete’ AOM and NOB. Based on kinetic theory of optimal pathway length^{5,6}, Costa *et al.*⁷ argued that a hypothetical complete nitrifier would likely be outcompeted by incomplete, cross-feeding AOM and NOB in many environments. However, the same authors⁷ also pointed out that a complete nitrifier might be competitive under conditions that favour the maximization of growth yield rather than growth rate and coined the term “comammox” (complete ammonia oxidizer) to describe such a hypothetical microbe. Conditions selecting for comammox may be characterized by slow, substrate-influx-limited growth with a spatial clustering of biomass in microbial aggregates and biofilms⁷. A prerequisite for the existence of comammox would also be that any biochemical incompatibilities of ammonia and nitrite oxidation can be overcome by adaptations of enzymes or cellular compartmentalization⁷. Aside from these theoretical considerations, the old question of whether comammox exists in nature has not been resolved.

The globally distributed genus *Nitrospira* represents the most diverse known group of NOB. *Nitrospira* members have been found in terrestrial⁸ and limnic habitats^{9,10}, marine waters¹¹, deep sea sediments, sponge tissue¹², geothermal springs¹³, drinking water distribution systems¹⁴, corroded iron pipes¹⁵, and wastewater treatment plants (WWTPs)^{10,16}. At least six phylogenetic sublineages of *Nitrospira* exist, of which lineage II seems to be most widely distributed in both natural and engineered ecosystems¹⁰. The ecological success of *Nitrospira* has been linked to an economical pathway for nitrite oxidation¹⁷ and a substantial metabolic versatility, which includes the utilization of various organic compounds in addition to nitrite and CO_2 ^{10,11,17–19}, cyanate or urea degradation and nitrification by reciprocal feeding with AOM^{19,20}, and chemolithoautotrophic aerobic hydrogen oxidation²¹.

Enrichment of conspicuous *Nitrospira*

A microbial biofilm developing on the walls of a pipe under the flow of hot water (56 °C, pH 7.5) raised from a 1,200 m deep oil exploration well (Aushiger, North Caucasus, Russia) was sampled and incubated at 46 °C in ammonium-containing mineral medium to enrich moderately thermophilic AOM. After a series of subcultivation steps, we obtained enrichment culture ‘ENR4’ that oxidized ammonia to nitrate and contained a dense population of cells morphologically resembling described *Nitrospira* species^{13,15} (Extended Data Fig. 1a, b). A second abundant population consisted of rod-shaped cells, but no organism in ENR4 displayed the typical morphologies of known AOM. Inspection by fluorescence *in situ* hybridization (FISH) with nitrifier-specific ribosomal RNA-targeted probes^{10,22} confirmed that ENR4 contained *Nitrospira* (Extended Data Fig. 1c) but no other detectable nitrifiers. Moreover, known bacterial or archaeal genes of ammonia monooxygenase (AMO) subunit alpha (*amoA*) and 16S rRNA genes of AOA were not detected by PCR in ENR4. Considering

¹Department of Microbiology and Ecosystem Science, Division of Microbial Ecology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria. ²Winogradsky Institute of Microbiology, Research Center of Biotechnology of the Russian Academy of Sciences, Leninsky Ave. 33, bld. 2, 119071 Moscow, Russia. ³Center for Microbial Communities, Department of Chemistry and Bioscience, Aalborg University, Fredrik Bajers Vej 7H, 9220 Aalborg, Denmark. ⁴Helmholtz-Centre for Environmental Research - UFZ, Department of Proteomics, Permoserstrasse 15, 04318 Leipzig, Germany. ⁵Helmholtz-Centre for Environmental Research - UFZ, Department of Metabolomics, Permoserstrasse 15, 04318 Leipzig, Germany. ⁶Department of Microbiology and Ecosystem Science, Division of Computational Systems Biology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria. ⁷DVGW-Forschungsstelle TUHH, Hamburg University of Technology, Am Schwarzenberg-Campus 3, 21073 Hamburg, Germany.

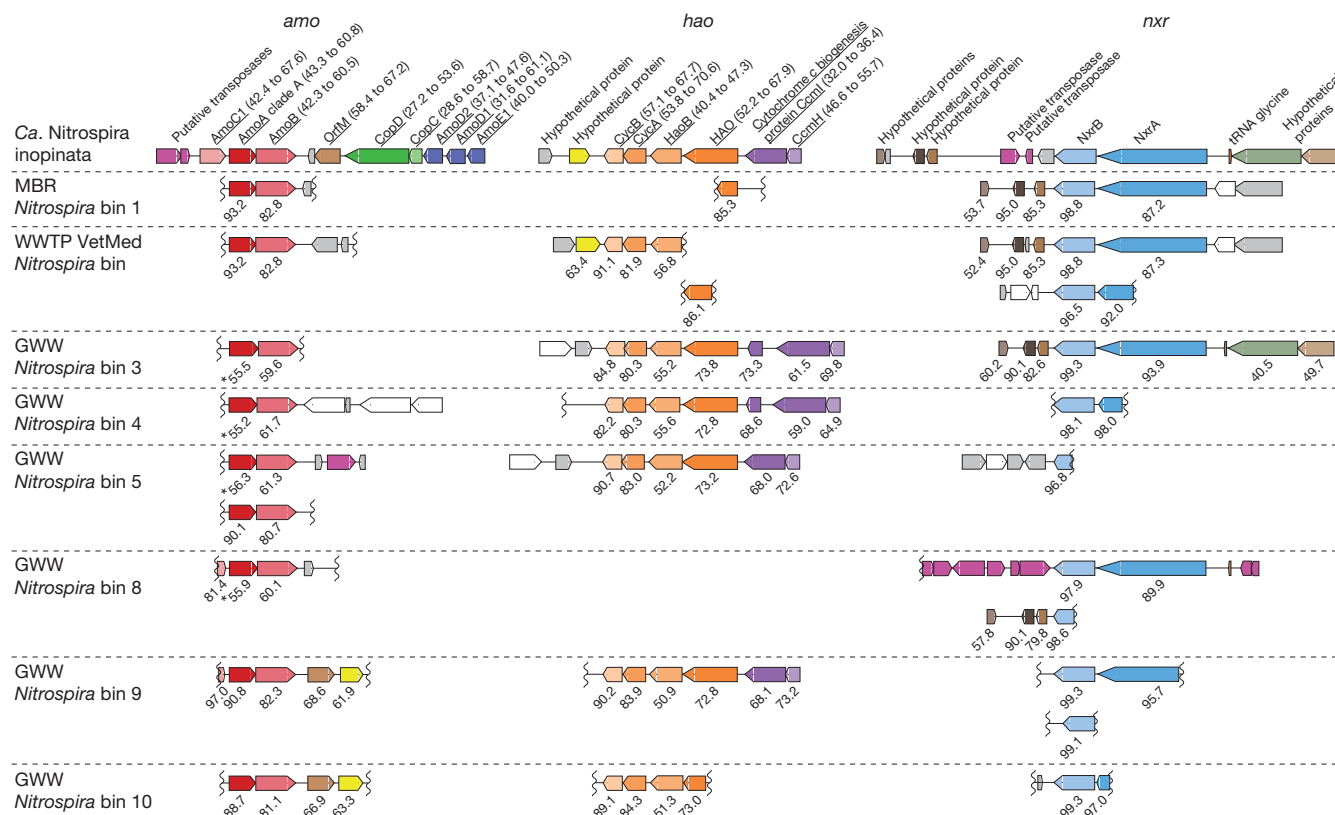


Figure 1 | Key nitrification gene loci in *Ca. N. inopinata* and the metagenomic *Nitrospira* population genome bins containing putative comammox *Nitrospira*. Gene alignments of the *amoCAB*, *hao*, and *nxrAB* loci with flanking genes are shown. Only two or three of up to nine syntenic cytochrome *c* biogenesis genes upstream of the *hao* loci are displayed. Colours identify homologous genes. Genes without homologues in the analysed data set are white if their function is known, otherwise grey. Transposases are magenta irrespectively of homology. Numbers below genes represent amino acid sequence identities (in per cent) of

the predicted gene products compared to *Ca. N. inopinata*. Asterisks mark comammox clade B *amoA* genes. Wiggly lines indicate ends of metagenomic contigs. Underlined gene products of *Ca. N. inopinata* have homologues in AOB genomes (amino acid identities in per cent to AOB are indicated in parentheses), but gene arrangements can differ from AOB²⁴. Genes and noncoding regions are drawn to scale. Metagenomic bins are numbered as in Supplementary Table 8. MBR, membrane bioreactor; WWTP, wastewater treatment plant; GWW, groundwater well.

the intriguing possibility that the *Nitrospira* population might be responsible for both ammonia and nitrite oxidation, we sequenced the metagenome of the enrichment (Supplementary Tables 1–7) to identify the ammonia oxidizer. Sequence assembly and differential coverage binning²³ showed that the ENR4 metagenome was dominated by two organisms (one *Nitrospira* strain and a betaproteobacterium affiliated with the family Hydrogenophilaceae) and revealed two additional rare populations (an alphaproteobacterium related to *Tepidamorphus gemmatus*, family Rhodobiaceae, and an actinobacterium affiliated with *Thermoleophilum*, family Thermoleophilaceae). Archaea were not detected (Extended Data Fig. 2a). Based on the relative genome sequence coverage in three sequenced samples of the culture, *Nitrospira* was the most abundant population in ENR4 (68 to 80% of the community), followed by the betaproteobacterium (18 to 29%) and the other two organisms ($\leq 2\%$). Subsequent FISH identified the relatively abundant rod-shaped cells as the betaproteobacterium (Extended Data Fig. 1c), whereas the two rare populations were encountered only sporadically by microscopy. Further subcultivation led to enrichment 'ENR6' that also oxidized ammonia to nitrate and, according to metagenome analysis, contained only *Nitrospira* (60% according to relative sequence coverage) and the betaproteobacterium (40%) (Extended Data Fig. 2b). The time of enrichment, from sampling of the source biofilm to ENR6, was four years. The high sequence coverage (Extended Data Fig. 2) allowed us to reconstruct complete and closed *Nitrospira* genomes and almost complete genomes of the other bacteria from the metagenomes of cultures ENR4 and ENR6, respectively. The *Nitrospira* genomes retrieved from the two enrichments were identical. We provisionally classify this highly enriched

Nitrospira strain as '*Candidatus Nitrospira inopinata*' (in.o.pi.na'ta. L. fem. adj. *inopinata* unexpected, surprising).

Discovery of comammox

The obtained bacterial genomes were screened for the key functional genes of autotrophic nitrification. As expected, *Ca. N. inopinata* possesses the key enzyme for nitrite oxidation, nitrite oxidoreductase (NXR). Its genome contains the *nxrA* and *nxrB* genes coding for the subunits alpha and beta, respectively, of the periplasmic *Nitrospira* NXR¹⁷ and genes of four candidate *Nitrospira* NxrC gamma subunits¹⁷ (Extended Data Fig. 3). Unlike other cultured *Nitrospira*, which possess two to five paralogous copies of the *nxrAB* genes^{8,17}, *Ca. N. inopinata* has only one copy of these genes. Much more surprisingly, *Ca. N. inopinata* also possesses homologues to the hallmark enzymes of ammonia oxidation, AMO and hydroxylamine dehydrogenase (also referred to as hydroxylamine oxidoreductase, HAO)²⁴ (Extended Data Fig. 1d). Its *amoA* gene is dissimilar to those of canonical AOM and was thus not picked up in the initial *amoA* PCR screening of ENR4. The three AMO subunits alpha (AmoA), beta (AmoB) and gamma (AmoC) are encoded by a single *amoCAB* gene cluster and by two additional *amoC* genes at other genomic loci (the AmoC copies share amino acid sequence identities of 99.63 to 100%) (Fig. 1, Extended Data Fig. 3). The *amoCAB* gene order is conserved in all AOB²⁴. *Ca. N. inopinata* also has homologues of the putative membrane-associated proteins AmoD and AmoE of AOB, which may interact with the ammonia-oxidizing machinery or the electron transport chain²⁵, and a homologue of the putative membrane protein OrfM found in all AOB²⁴ (Fig. 1, Extended Data Fig. 3). Similar to betaproteobacterial

AOB²⁴, genes of the copper resistance proteins CopC/D are located close to the *amo* locus (Fig. 1). The single *hao* gene of *Ca. N. inopinata* encodes a predicted octahem cytochrome *c* protein resembling the HAO of AOB²⁶. Like in AOB, *hao* shares a genomic locus with gene *haoB* of a putative membrane protein found in all AOB²⁷ and with two genes of tetrahaem *c*-type cytochromes, which resemble cytochrome *c*₅₅₄ (CycA) and cytochrome *c*_{m552} (CycB) of AOB²⁴ (Fig. 1). HAO, CycA, and CycB form the hydroxylamine ubiquinone reduction module (HURM) in AOB, which transfers electrons from hydroxylamine to the quinone pool²⁸. The full genetic complement for both ammonia and nitrite oxidation strongly suggested that *Ca. N. inopinata* is a comammox organism (Extended Data Fig. 1d). No canonical nitrification genes were found in the genomes of the other three bacteria detected in ENR4, suggesting that these co-enriched organisms were heterotrophs that used organic substrates produced by the autotrophic *Nitrospira*²⁹. The betaproteobacterial genome, which was identical in enrichments ENR4 and ENR6, encodes a membrane-associated nitrate reductase that is highly similar to the known nitrate reductases of *E. coli* and other Proteobacteria.

Phylogenetic inference based on 16S rRNA gene sequences showed that *Ca. N. inopinata* belongs to the widely distributed lineage II of the genus *Nitrospira*¹⁰ (Extended Data Fig. 4). The other cultured *Nitrospira* strains in lineage II are *N. moscoviensis*¹⁵, *N. lenta*³⁰, and *N. japonica*³¹, which are NOB and do not possess the enzymatic repertoire to utilize ammonia as energy source. Consistently, the affiliation of *Ca. N. inopinata* with *Nitrospira* lineage II was supported by phylogenies based on *nrrB* gene and NxrA protein sequences (Extended Data Fig. 5). The *nrrB* gene is a suitable phylogenetic marker to differentiate the *Nitrospira* lineages⁸. NxrA phylogenies reliably distinguish *Nitrospira* NxrA from related enzymes¹⁷, but their resolution within the genus *Nitrospira* has not been evaluated and assignments of NxrA sequences to specific *Nitrospira* lineages must be treated with caution. *Ca. N. inopinata* represents a different species from the two comammox *Nitrospira* strains described by van Kessel *et al.*³², on the basis of the low pairwise average nucleotide identities (70.3 to 71.6%) between the genomes of *Ca. N. inopinata* and these organisms.

Full nitrification by *Ca. N. inopinata*

Complete nitrification by *Ca. N. inopinata* was demonstrated by incubation experiments in mineral media containing ammonium as the sole source of energy and reductant, and bicarbonate/CO₂ as the sole carbon source. Consistent with the anticipated activity of comammox, *Ca. N. inopinata* nearly stoichiometrically oxidized 1 mM or 0.1 mM ammonium to nitrate (Fig. 2a, b). A transient accumulation of nitrite (up to 30% of the added ammonium) was observed in parallel to nitrate production, but nitrite was completely oxidized after all ammonium had been consumed. Much lower nitrite accumulation was detected in an experiment with 10 μ M ammonium (Fig. 2c), suggesting that experimental parameters strongly influence this phenomenon and that nitrite accumulation might actually not occur under environmentally relevant conditions. Apparent nitrogen loss caused by formation of gaseous compounds was not observed in any experiment (Fig. 2), suggesting that NO formation from nitrite by NirK (Extended Data Fig. 1d) was not quantitatively important for *Ca. N. inopinata*. Growth of *Ca. N. inopinata* during oxidation of ammonium to nitrate was demonstrated by quantitative PCR targeting its single-copy *amoA* gene and continued after consumption of ammonium in the presence of nitrite until the end of the experiment (Fig. 2d). A pure culture of the betaproteobacterium, the only non-*Nitrospira* microbe in ENR6, was isolated in acetate-containing medium and showed no nitrifying activity after inoculation into ammonium- or nitrite-media at cell densities higher than the density of *Ca. N. inopinata* in the growth experiment (Fig. 2d, Extended Data Fig. 6). The function of *Ca. N. inopinata* as comammox was further confirmed by metaproteome analysis of ENR4, which showed that all key proteins of *Ca. N. inopinata* for ammonia and nitrite oxidation were expressed during incubation with

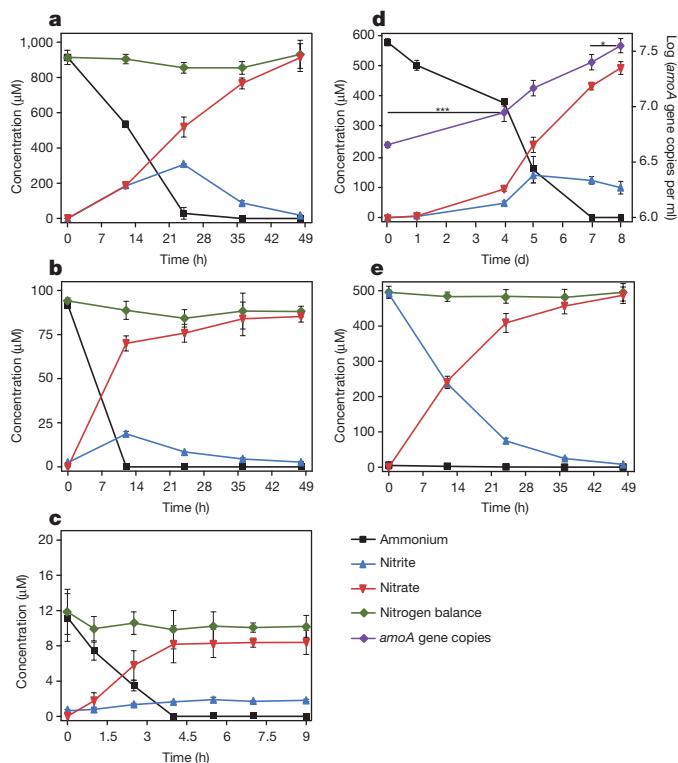


Figure 2 | Complete nitrification by *Ca. N. inopinata* in enrichment culture ENR4. a–c, Near-stoichiometric oxidation of 1 mM, 0.1 mM, or 10 μ M ammonium to nitrate with transient accumulation of nitrite. d, Growth of *Ca. N. inopinata* on ammonium (initial concentration 0.6 mM) as determined by qPCR of the *amoA* gene. Ammonia oxidation was slow because this experiment was started with a highly diluted culture. Significance of difference was calculated by a paired *t*-test (* $P < 0.05$; *** $P < 0.01$) between data points as indicated by horizontal lines. e, Near-stoichiometric oxidation of 0.5 mM nitrite to nitrate by ammonium-grown *Ca. N. inopinata*. The cells were washed to remove residual ammonium before nitrite addition. Data points show means, error bars show 1 s.d. of $n = 4$ (a, b, e) or $n = 3$ (c, d) biological replicates. If not visible, error bars are smaller than symbols.

ammonium and that NXR, HAO and AmoB were among the 50 most abundant proteins of this organism (Extended Data Fig. 7).

When a culture grown on ammonium was transferred into mineral medium containing only 0.5 mM nitrite as energy source and electron donor, *Ca. N. inopinata* stoichiometrically oxidized nitrite to nitrate (Fig. 2e). However, subsequent additions of nitrite did not result in further nitrite oxidation. We hypothesize that nitrite was first oxidized by residual NXR activity, but metabolic activity and biosynthetic processes finally stalled in the absence of ammonium. This would be consistent with an incapability of *Ca. N. inopinata* to use nitrite as nitrogen source for assimilation due to the absence of genes for a nitrite transporter and assimilatory nitrite reductase. It is interesting to note that *Ca. N. inopinata* could theoretically utilize nitrite as nitrogen source by respiratory ammonification catalysed by a periplasmic cytochrome *c* nitrite reductase (NrfA) using electrons from quinol (Extended Data Fig. 1d). Genes for respiratory ammonification have not been detected in the genomes of the two comammox strains reported by van Kessel *et al.*³² and of other *Nitrospira*^{17,19}. However, the *Ca. N. inopinata* genome lacks a second copy of respiratory complex III, which may be needed for the reverse transport of high-potential electrons from nitrite to quinone in other *Nitrospira*¹⁷. If nitrite is the only available electron donor, this gap in the reverse electron transport chain probably prevents nitrite reduction by NrfA and CO₂ fixation by the reductive tricarboxylic acid cycle, which is the autotrophic pathway in *Ca. N. inopinata* (Extended Data Fig. 1d) and other *Nitrospira*¹⁷. Thus, *Ca. N. inopinata* may grow on nitrite only in the presence of

a low-potential electron donor for quinone reduction such as H_2 (ref. 21) or intracellular storage compounds (Extended Data Fig. 1d). In contrast, electrons derived from ammonia should be transferred to quinone as in AOB¹ and thus enable autotrophic growth (Extended Data Fig. 1d). An ammonium transporter also enables *Ca. N. inopinata* to use ammonium as nitrogen source. The incapability to grow on nitrite distinguishes *Ca. N. inopinata* from the strictly nitrite-oxidizing *Nitrospira* that grow on nitrite and CO_2 ^{15,17}.

Distinct AMO and HAO of *Ca. N. inopinata*

The *amoA* gene is a functional and phylogenetic marker for AOM^{33,34}, which has been used in numerous studies as a cultivation-independent tool to detect and identify AOM in microbial communities. Intriguingly, phylogenetic analyses revealed that *Ca. N. inopinata* possesses a new type of AmoA that differs from the AmoA forms of known AOB and AOA. It belongs to a distinct clade ('comammox AmoA clade A') that contains numerous environmental sequences and shares a common ancestor with the AmoA lineage of the betaproteobacterial AOB (Fig. 3, Extended Data Fig. 8). Similar to the AmoA phylogeny, the *amoB* and *amoC* as well as the *hao* genes of *Ca. N. inopinata* fell into distinct lineages that are related to the respective homologues of AOB (Extended Data Fig. 9).

A fascinating question is whether the unusual enzymes for ammonia oxidation are ancestral features of the genus *Nitrospira* or were acquired by *Ca. N. inopinata* by lateral gene transfer. The first scenario would imply that these genes have been lost by the strictly nitrite-oxidizing *Nitrospira* members. Indications for a possible lateral gene transfer event in *Ca. N. inopinata* are putative transposase genes directly upstream of the *amoCAB* genes (Fig. 1) and a tetranucleotide pattern of the *amoCAB*-containing region that significantly deviates from the genome-wide signature (Extended Data Fig. 10). The tetranucleotide pattern of the *amoCAB* region also clearly differs from the genome-wide signature of the betaproteobacterium found in ENR4 and ENR6, strongly suggesting that these genes did not originate from this strain. Putative transposases are also located downstream of the *nxxAB* operon, whose tetranucleotide pattern also deviates from the genome-wide signature (Fig. 1, Extended Data Fig. 10). *Ca. N. inopinata* belongs to *Nitrospira* lineage II (Extended Data Fig. 4), and its NXR is also affiliated with lineage II (Extended Data Fig. 5). Thus, if lateral gene transfer occurred, the *nxx* genes must have been received from another *Nitrospira* lineage II member. No indications were found for lateral gene transfer of the other two *amoC* copies or of *hao* and the other HURM genes of *Ca. N. inopinata* (Fig. 1, Extended Data Fig. 10).

Distribution of comammox *Nitrospira*

A screening of public databases retrieved sequences within comammox AmoA clade A, which originated from paddy and other agricultural soils, forest soils, paddy field floodwater, freshwater environments such as wetlands, river beds, aquifers and lake sediments, and from engineered systems (activated sludge and drinking water treatment plants) (Extended Data Fig. 8). For most of these sequences no quantitative information regarding their abundance is available. However, for three metagenomic data sets from Rifle soils³⁵, relative abundances could be estimated from raw sequence data. In these soils, archaeal *amoA* sequences were found to be 3.8- to 10.5-fold more abundant than comammox *amoA* sequences. Interestingly, only very low numbers of betaproteobacterial *amoA* sequences were found and those were retrieved exclusively from the unassembled Rifle data sets. Additional database searches retrieved sequences from soil, freshwater and engineered environments that clustered in phylogenetic trees with the *amoB*, *amoC* and *hao* genes of *Ca. N. inopinata* (Extended Data Fig. 9). These results are consistent with a wide environmental distribution of comammox organisms with the possible exception of oceanic environments, as no comammox marker genes were identified in marine metagenomes.

To further elucidate the distribution of putative comammox *Nitrospira* in engineered systems, we sequenced total community

metagenomes from a pilot-scale membrane bioreactor (MBR) at the municipal WWTP Aalborg West (Denmark) and of nitrifying activated sludge from the full-scale WWTP of the University of Veterinary Medicine, Vienna, Austria ('WWTP VetMed'). A great diversity of lineage I and II *Nitrospira* had previously been detected in WWTP VetMed¹⁸. Quantitative FISH in the VetMed sample used for metagenomics revealed dominance of *Nitrospira* ($7.5 \pm 3\%$ of all detectable bacteria, 1 s.d.) over AOB ($2.5 \pm 1.2\%$). AOA did not occur in this WWTP at an abundance relevant for nitrification as no sequences affiliated with *Thaumarchaeota* were detected in the metagenomic data set. Additionally, we sequenced metagenomes of pasty and suspended iron sludge from a groundwater well (GWW) of a waterworks (Wolfenbüttel, Germany) (Supplementary Tables 1 and 2). *Nitrospira* population genome bins were retrieved by differential coverage binning²³ from all metagenomes (Supplementary Table 8). According to 16S rRNA and NXR phylogenies, these *Nitrospira* belonged to lineages I and II (Extended Data Figs 4 and 5). Intriguingly, *amo* and *hao* genes similar to those of *Ca. N. inopinata* were found in one or more *Nitrospira* bins from every sample, suggesting that comammox *Nitrospira* frequently occur in engineered systems (Extended Data Figs 8 and 9, Fig. 1). However, not all *Nitrospira* bins contained genes for ammonia oxidation (Supplementary Table 8). Since *nxx* genes were found in all bins except GWW *Nitrospira* bin 7, we assume that comammox coexisted in these communities with *Nitrospira* that were strict nitrite oxidizers or used alternate metabolisms^{19,21}. In several *Nitrospira* bins with sufficiently long contigs, the *amo*, *hao* and *nxx* loci and flanking genes were syntenic with *Ca. N. inopinata* (Fig. 1). In particular, a gene cluster for cytochrome *c* biogenesis is located directly upstream of the *hao* gene in *Ca. N. inopinata* and metagenomes (Fig. 1). This gene arrangement is not found in genome-sequenced AOB^{36–40}, suggesting that it may be diagnostic for comammox *Nitrospira*. Transposases were found close to the *amo* and *nxx* genes, respectively, in two *Nitrospira* bins (Fig. 1). Moreover, a second type of AmoA was identified in some of the GWW *Nitrospira* bins. These sequences fell into a phylogenetic sister lineage of comammox AmoA clade A, which also contains other environmental sequences from soil and freshwater ecosystems (Extended Data Fig. 8, Fig. 3), and showed considerably lower identities to the AmoA of *Ca. N. inopinata* (Fig. 1). Consequently, we refer to this lineage as 'comammox AmoA clade B'. The *amoB* and *amoC* sequences from those *Nitrospira* bins, which contained clade B-*amoA*, also fell separately in the respective phylogenetic trees (Extended Data Fig. 9a, b). Thus, two different and related new types of AMO occur in bacteria from the genus *Nitrospira*, and both share a common ancestor with the AMO of the betaproteobacterial AOB (Fig. 3).

We have noticed that sequences in comammox AmoA clades A and B were previously assigned as particulate methane monooxygenase (PmoA) to uncultured gammaproteobacterial (clade A, *Crenothrix polyspora*) and alphaproteobacterial (clade B) methanotrophs^{41,42}. While these assignments were based on indirect evidence, our study provides direct physiological proof that an organism expressing an enzyme in clade A oxidizes ammonia and metagenomic evidence for a *Nitrospira* origin of genes in both clades. However, it remains possible that genes in these clades were exchanged by lateral gene transfer between nitrifiers and methanotrophs.

Discussion

The first cultured comammox organism *Ca. N. inopinata* is a moderately thermophilic *Nitrospira* member, and uncultured mesophilic comammox *Nitrospira* were identified by metagenomics in this study, too. The genus *Nitrospira* is one of the most diverse^{8,18} known nitrifier groups and colonizes virtually all oxic ecosystems¹⁰, including high-temperature environments^{13,15}. It is tempting to speculate that the environmental distribution of comammox is largely congruent with that of *Nitrospira*, which are mostly uncultured and poorly characterized. Previous research was based on the dogma that all

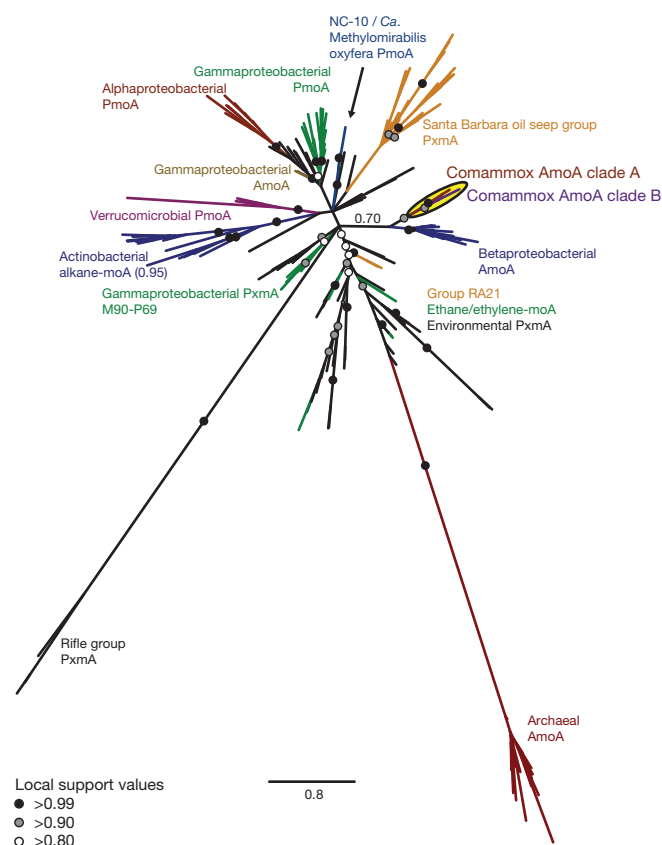


Figure 3 | Phylogenetic affiliation of comammox AmoA sequences to other AmoA superfamily members. Bayesian inference tree showing the phylogenetic relationship of comammox AmoA to other members of the AmoA superfamily (202 taxa, 238 alignment positions). Comammox AmoA sequences formed clades A (posterior probability, PP = 0.99) and B (PP = 0.97) that grouped together (PP = 0.91) and with betaproteobacterial AmoA (PP = 0.70). Scale bar indicates estimated change per nucleotide.

Nitrospira use nitrite, but not ammonia, as an energy source. Due to this firm expectation, comammox *Nitrospira* were overlooked for decades and some repeatedly observed phenomena could not be well explained. For example, conspicuously high *in situ* abundances of uncultured *Nitrospira*, which exceeded the abundances of known AOM in the same samples, were detected in nitrifying biofilms, activated sludge, freshwater sediments, and drinking water distribution systems^{14,18,43–45}. These puzzling observations are inconsistent with the classical concept of nitrification, which suggests an AOM:NOB ratio greater than one⁴⁶. Aside from other energy-conserving metabolic activities of NOB in addition to nitrite oxidation^{10,21,46}, the presence of comammox organisms in those *Nitrospira* communities would be a plausible explanation for the increased ratio of *Nitrospira* over known AOM. Indeed, we detected *amo* and *hao* genes in the *Nitrospira* metagenome from WWTP VetMed (Fig. 1, Extended Data Figs 8 and 9), a system in which *Nitrospira* outnumber AOB according to FISH and comammox represents 43 to 71% of the *Nitrospira* population as estimated from gene abundances in the metagenomic data sets. A high relative abundance of comammox (58 to 74% of all *Nitrospira*) was also estimated for the GWW based on metagenome analysis. More precise analyses of comammox abundance as well as its spatial interactions with other community members will require the development of assays to rapidly differentiate *in situ* between strictly nitrite-oxidizing and comammox *Nitrospira*.

Studies with strictly nitrite-oxidizing representatives of this genus characterized *Nitrospira* as slow-growing microbial K-strategists adapted to low substrate concentrations^{18,43,47,48}. Many *Nitrospira*, including *Ca. N. inopinata*, also form microcolonies, flocs, and biofilms^{10,43}. These properties, if generally shared by comammox

Nitrospira, would be in agreement with the theoretically predicted⁷ ecological niche of comammox. The engineered systems surveyed in this study are characterized by biofilm or floc formation. Diffusion barriers and ammonium or nitrite concentration gradients⁴⁷ in biofilms could create niches with limited substrate influx, where comammox might outcompete incomplete nitrifiers. Complex biofilm or floc architectures with numerous microenvironments may support diverse nitrifier communities like in WWTP VetMed, which consist of comammox as well as canonical AOB and NOB. Future comammox isolates from the *Ca. N. inopinata* culture and from other enrichments may offer chances to experimentally define the conditions that select for these organisms and to study the competition of comammox with other nitrifiers, including strictly nitrite-oxidizing *Nitrospira* and AOA adapted to low substrate concentrations^{48,49}.

The discovery of comammox has revealed that the division of metabolic labour in nitrification is not obligate and will thus have far-reaching implications for future studies on the microbiology of nitrogen cycling. It opens a new field in nitrification research and some of the most pressing open questions range from the biochemistry, regulation, inhibition, and kinetics of complete nitrification to the diversity, population dynamics, metabolic versatility, and biological interactions of comammox organisms. In particular, the integration of comammox in studies on the niche specialization and niche partitioning of AOB and AOA⁵⁰ or NOB⁴³ will be crucial to obtain a picture of nitrification as it actually occurs in nature. Such insights may lead to refined strategies to manage nitrification in sewage treatment, drinking water supply, and agriculture. The presence of new AMO and HAO types, which share common ancestry with these enzymes of betaproteobacterial AOB, in the phylogenetically deep-branching genus *Nitrospira*¹⁵ impressively exemplifies the modular evolution of the nitrogen cycle²⁸ and adds further complexity to the intricate evolutionary history of nitrification^{17,28}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 August; accepted 19 November 2015.

Published online 26 November 2015.

- Bock, E. & Wagner, M. in *The Prokaryotes: A Handbook on the Biology of Bacteria* (eds Dworkin, M. et al.) 457–495 (Springer, 2001).
- Könneke, M. et al. Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**, 543–546 (2005).
- Winogradsky, S. Contributions a la morphologie des organismes de la nitrification. *Arch. Sci. Biol. (St. Petersburg)* **1**, 87–137 (1892).
- Arp, D. & Bottomley, P. J. Nitrifiers: more than 100 years from isolation to genome sequences. *Microbe* **1**, 229–234 (2006).
- Pfeiffer, T. & Bonhoeffer, S. Evolution of cross-feeding in microbial populations. *Am. Nat.* **163**, E126–E135 (2004).
- Heinrich, R. & Schuster, S. *The regulation of cellular systems*. (Chapman & Hall, 1996).
- Costa, E., Pérez, J. & Kreft, J. U. Why is metabolic labour divided in nitrification? *Trends Microbiol.* **14**, 213–219 (2006).
- Pester, M. et al. *NxrB* encoding the beta subunit of nitrite oxidoreductase as functional and phylogenetic marker for nitrite-oxidizing *Nitrospira*. *Environ. Microbiol.* **16**, 3055–3071 (2014).
- Hovanec, T. A., Taylor, L. T., Blakis, A. & DeLong, E. F. *Nitrospira*-like bacteria associated with nitrite oxidation in freshwater aquaria. *Appl. Environ. Microbiol.* **64**, 258–264 (1998).
- Daims, H., Nielsen, J. L., Nielsen, P. H., Schleifer, K. H. & Wagner, M. In situ characterization of *Nitrospira*-like nitrite-oxidizing bacteria active in wastewater treatment plants. *Appl. Environ. Microbiol.* **67**, 5273–5284 (2001).
- Watson, S. W., Bock, E., Valois, F. W., Waterbury, J. B. & Schlosser, U. *Nitrospira marina* gen. nov. sp. nov.: a chemolithotrophic nitrite-oxidizing bacterium. *Arch. Microbiol.* **144**, 1–7 (1986).
- Taylor, M. W., Radax, R., Steger, D. & Wagner, M. Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol. Mol. Biol. Rev.* **71**, 295–347 (2007).
- Lebedeva, E. V. et al. Isolation and characterization of a moderately thermophilic nitrite-oxidizing bacterium from a geothermal spring. *FEMS Microbiol. Ecol.* **75**, 195–204 (2011).
- Martiny, A. C., Albrechtsen, H. J., Arvin, E. & Molin, S. Identification of bacteria in biofilm and bulk water samples from a nonchlorinated model drinking water distribution system: detection of a large nitrite-oxidizing population associated with *Nitrospira* spp. *Appl. Environ. Microbiol.* **71**, 8611–8617 (2005).

15. Ehrlich, S., Behrens, D., Lebedeva, E., Ludwig, W. & Bock, E. A new obligately chemolithoautotrophic, nitrite-oxidizing bacterium, *Nitrospira moscoviensis* sp. nov. and its phylogenetic relationship. *Arch. Microbiol.* **164**, 16–23 (1995).
16. Schramm, A., De Beer, D., Wagner, M. & Amann, R. Identification and activities in situ of *Nitrosospora* and *Nitrospira* spp. as dominant populations in a nitrifying fluidized bed reactor. *Appl. Environ. Microbiol.* **64**, 3480–3485 (1998).
17. Lückner, S. *et al.* A *Nitrospira* metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proc. Natl Acad. Sci. USA* **107**, 13479–13484 (2010).
18. Gruber-Dorninger, C. *et al.* Functionally relevant diversity of closely related *Nitrospira* in activated sludge. *ISME J.* **9**, 643–655 (2015).
19. Koch, H. *et al.* Expanded metabolic versatility of ubiquitous nitrite-oxidizing bacteria from the genus *Nitrospira*. *Proc. Natl Acad. Sci. USA* **112**, 11371–11376 (2015).
20. Palatinszky, M. *et al.* Cyanate as an energy source for nitrifiers. *Nature* **524**, 105–108 (2015).
21. Koch, H. *et al.* Growth of nitrite-oxidizing bacteria by aerobic hydrogen oxidation. *Science* **345**, 1052–1054 (2014).
22. Mobarry, B. K., Wagner, M., Urbain, V., Rittmann, B. E. & Stahl, D. A. Phylogenetic probes for analyzing abundance and spatial organization of nitrifying bacteria. *Appl. Environ. Microbiol.* **62**, 2156–2162 (1996).
23. Albertsen, M. *et al.* Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature Biotechnol.* **31**, 533–538 (2013).
24. Arp, D. J., Chain, P. S. G. & Klotz, M. G. The impact of genome analyses on our understanding of ammonia-oxidizing bacteria. *Annu. Rev. Microbiol.* **61**, 503–528 (2007).
25. Berube, P. M. & Stahl, D. A. The divergent AmoC₃ subunit of ammonia monooxygenase functions as part of a stress response system in *Nitrosomonas europaea*. *J. Bacteriol.* **194**, 3448–3456 (2012).
26. Klotz, M. G. *et al.* Evolution of an octahem cytochrome c protein family that is key to aerobic and anaerobic ammonia oxidation by bacteria. *Environ. Microbiol.* **10**, 3150–3163 (2008).
27. Bergmann, D. J., Hooper, A. B. & Klotz, M. G. Structure and sequence conservation of *hao* cluster genes of autotrophic ammonia-oxidizing bacteria: evidence for their evolutionary history. *Appl. Environ. Microbiol.* **71**, 5371–5382 (2005).
28. Klotz, M. G. & Stein, L. Y. Nitrifier genomics and evolution of the nitrogen cycle. *FEMS Microbiol. Lett.* **278**, 146–156 (2008).
29. Rittmann, B. E., Regan, J. M. & Stahl, D. A. Nitrification as a source of soluble organic substrate in biological treatment. *Water Sci. Technol.* **30**, 1–8 (1994).
30. Nowka, B., Off, S., Daims, H. & Spieck, E. Improved isolation strategies allowed the phenotypic differentiation of two *Nitrospira* strains from widespread phylogenetic lineages. *FEMS Microbiol. Ecol.* **91**, fiu031 (2015).
31. Ushiki, N., Fujitani, H., Aoi, Y. & Tsuneda, S. Isolation of *Nitrospira* belonging to sublineage II from a wastewater treatment plant. *Microbes Environ.* **28**, 346–353 (2013).
32. van Kessel, M. A. H. J. *et al.* Complete nitrification by a single microorganism. *Nature* <http://dx.doi.org/10.1038/nature16459> (2015).
33. Pester, M. *et al.* *amoA*-based consensus phylogeny of ammonia-oxidizing archaea and deep sequencing of *amoA* genes from soils of four different geographic regions. *Environ. Microbiol.* **14**, 525–539 (2012).
34. Purkhold, U. *et al.* Phylogeny of all recognized species of ammonia oxidizers based on comparative 16S rRNA and *amoA* sequence analysis: implications for molecular diversity surveys. *Appl. Environ. Microbiol.* **66**, 5368–5382 (2000).
35. Wrighton, K. C. *et al.* Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* **337**, 1661–1665 (2012).
36. Stein, L. Y. *et al.* Whole-genome analysis of the ammonia-oxidizing bacterium, *Nitrosomonas eutropha* C91: implications for niche adaptation. *Environ. Microbiol.* **9**, 2993–3007 (2007).
37. Chain, P. *et al.* Complete genome sequence of the ammonia-oxidizing bacterium and obligate chemolithoautotroph *Nitrosomonas europaea*. *J. Bacteriol.* **185**, 2759–2773 (2003).
38. Suwa, Y. *et al.* Genome sequence of *Nitrosomonas* sp. strain AL212, an ammonia-oxidizing bacterium sensitive to high levels of ammonia. *J. Bacteriol.* **193**, 5047–5048 (2011).
39. Norton, J. M. *et al.* Complete genome sequence of *Nitrosospora multiformis*, an ammonia-oxidizing bacterium from the soil environment. *Appl. Environ. Microbiol.* **74**, 3559–3572 (2008).
40. Klotz, M. G. *et al.* Complete genome sequence of the marine, chemolithoautotrophic, ammonia-oxidizing bacterium *Nitrosococcus oceanus* ATCC 19707. *Appl. Environ. Microbiol.* **72**, 6299–6315 (2006).
41. Radajewski, S. *et al.* Identification of active methylotroph populations in an acidic forest soil by stable-isotope probing. *Microbiology* **148**, 2331–2342 (2002).
42. Stoecker, K. *et al.* Cohn's *Crenothrix* is a filamentous methane oxidizer with an unusual methane monooxygenase. *Proc. Natl Acad. Sci. USA* **103**, 2363–2367 (2006).
43. Schramm, A., de Beer, D., van den Heuvel, J. C., Ottengraf, S. & Amann, R. Microscale distribution of populations and activities of *Nitrosospora* and *Nitrospira* spp. along a macroscale gradient in a nitrifying bioreactor: quantification by in situ hybridization and the use of microsensors. *Appl. Environ. Microbiol.* **65**, 3690–3696 (1999).
44. Altmann, D., Stief, P., Amann, R., De Beer, D. & Schramm, A. In situ distribution and activity of nitrifying bacteria in freshwater sediment. *Environ. Microbiol.* **5**, 798–803 (2003).
45. Foessel, B. U. *et al.* *Nitrosomonas* Nm143-like ammonia oxidizers and *Nitrospira marina*-like nitrite oxidizers dominate the nitrifier community in a marine aquaculture biofilm. *FEMS Microbiol. Ecol.* **63**, 192–204 (2008).
46. Winkler, M. K. H., Bassin, J. P., Kleerebezem, R., Sorokin, D. Y. & van Loosdrecht, M. C. M. Unravelling the reasons for disproportion in the ratio of AOB and NOB in aerobic granular sludge. *Appl. Microbiol. Biotechnol.* **94**, 1657–1666 (2012).
47. Maixner, F. *et al.* Nitrite concentration influences the population structure of *Nitrospira*-like bacteria. *Environ. Microbiol.* **8**, 1487–1495 (2006).
48. Nowka, B., Daims, H. & Spieck, E. Comparison of oxidation kinetics of nitrite-oxidizing bacteria: nitrite availability as a key factor in niche differentiation. *Appl. Environ. Microbiol.* **81**, 745–753 (2015).
49. Martens-Habbena, W., Berube, P. M., Urakawa, H., de la Torre, J. R. & Stahl, D. A. Ammonia oxidation kinetics determine niche separation of nitrifying *Archaea* and *Bacteria*. *Nature* **461**, 976–979 (2009).
50. Prosser, J. I. & Nicol, G. W. Archaeal and bacterial ammonia-oxidisers in soil: the quest for niche specialisation and differentiation. *Trends Microbiol.* **20**, 523–531 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank T. K. Lee and M. Steinberger for help with PCR analyses, N. V. Grigor'eva and M. Pogoda for assistance with culture maintenance, N. A. Kostrikina for assistance with electron microscopy, K. Kitzinger for support with FISH analyses, M. Mooshammer for help with chemical analyses, R. Hatzenpichler for designing probe Nmr1009, K. Eismann for help with proteome sample preparation, B. Scheer for help with mass spectrometer maintenance, Puren GmbH (Wolfenbüttel, Germany) for cooperation, N. Chernykh and J. Rosenthal for taking samples, and H. Koch and E. Bock for discussion. The authors are grateful for using the analytical facilities of the Centre for Chemical Microscopy (ProVIS) (Helmholtz Centre for Environmental Research), which is headed by H. Richnow (Department of Isotope Biochemistry) and supported by European Regional Development Funds (EFRE–Europe funds Saxony) and the Helmholtz Association. P.P. and H.D. were supported by the Austrian Science Fund (FWF) projects P27319-B21 and P25231-B21 (to H.D.). M.P., J.V., P.H., and M.W. were supported by the European Research Council Advanced Grant project NITRICARE 294343 (to M.W.). M.A., R.H.K., and P.H.N. were supported by the Danish Council for Independent Research, DFF – 4005-00369 and Innovation Fund Denmark (EcoDesign).

Author Contributions H.D. did (meta)genomic analysis of *Ca. N. inopinata* and comammox *Nitrospira*, contributed to phylogenetic and proteomics data analyses, designed the study and wrote the paper; E.V.L. enriched *Ca. N. inopinata*; E.V.L., P.P., P.H., A.B. and M.P. performed physiological experiments, analysed data, and characterized enrichments; M.A., R.H.K. and P.H.N. carried out metagenome sequencing, assembly and binning; C.H. performed bioinformatic and phylogenetic analyses; N.J. and M.v.B. performed proteomics measurements and data analysis; T.R. performed bioinformatic analyses; P.H., M.P. and J.V. maintained enrichment cultures and performed experiments; J.V. carried out database analyses; B.B. organized sampling and characterized environmental samples; M.W. designed the study, analysed data, and wrote the paper. All authors discussed the results and commented the manuscript.

Author Information All raw sequence data is available in the European Nucleotide Archive (ENA) under the project accession number PRJEB10139. The genome sequence of *Ca. N. inopinata* has been deposited at ENA under the project PRJEB10818, sequence accession LN885086. The draft genome of the betaproteobacterium from ENR4 and ENR6 is available in the JGI Integrated Microbial Genomes Database (<https://img.jgi.doe.gov/cgi-bin/m/main.cgi>) under the IMG Genome ID 2636415980. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.W. (wagner@microbial-ecology.net).

METHODS

Sampling sites. The inoculum for the *Ca. N. inopinata* enrichment culture was sampled from a microbial biofilm that grew on the metal surface of a pipe and was covered by hot water, which was raised from a 1,200 m deep oil exploration well. The water temperature was 56 °C and the pH 7.5. The well was located in Aushiger, North Caucasus, Russia (43°22′45.0″ N, 43°43′26.1″ E). The biofilm samples were taken in April 2011. Activated sludge, membrane biofilm, and foam (from a foaming event) samples were taken in August and October 2014 from a pilot-scale membrane bioreactor (MBR) performing nitrogen removal and enhanced biological phosphorus removal (EBPR) at the conventional full-scale WWTP Aalborg West, Aalborg, Denmark (57°02′59.9″ N, 9°51′55.4″ E). The influent wastewater for this MBR came from the primary settling tank of the full-scale plant, entering an anoxic/denitrification (2 m³) tank and going to an oxic/nitrification (2 m³) tank. An anaerobic tank (1.8 m³) used for return sludge side-stream hydrolysis provided easily degradable substrate for EBPR and denitrification. Activated sludge was also sampled from an aerated activated sludge basin (tank no. 2) of the full-scale WWTP of the University of Veterinary Medicine, Vienna, Austria (48°15′17.8″ N, 16°25′45.6″ E) in January 2015 (WWTP VetMed). The two continuously operated activated sludge tanks of this WWTP have a volume of 254 m³ each. The wastewater composition and nitrogen load vary with the amounts of animal faeces and other sewage. This WWTP was known to host a large diversity of *Nitrospira*¹⁸. Iron sludge samples were taken from groundwater well (GWV) no. 1 of the well field of the Wolfenbüttel waterworks (Wolfenbüttel, Germany) (52°08′55.9″ N, 10°32′33.9″ E). The well has a depth of 50 m below ground level (bgl) and a diameter of 600 mm. Groundwater is extracted through two well intake screens in 28 to 38 m bgl and 46 to 48 m bgl. The normal well capacity is 160 m³ h⁻¹. Before sampling, the well had been out of operation for about three weeks. The well water is a mixture of aerobic and anaerobic groundwater from two different ground water storeys and is characterized by the following parameters (values from years 2012 to 2014): pH about 7.2, about 10 °C, 5 to 10 mg l⁻¹ dissolved oxygen, 0.13 to 0.17 mg l⁻¹ ammonium, <0.01 mg l⁻¹ nitrite, 12 to 16 mg l⁻¹ nitrate, 0.16 to 0.42 mg l⁻¹ total iron, 0.03 to 0.08 mg l⁻¹ manganese, 0.64 to 0.99 mg l⁻¹ total organic carbon, 0.44 to 0.78 mg l⁻¹ dissolved organic carbon, 71 to 81 mg l⁻¹ dissolved inorganic carbon, 121 to 138 mg l⁻¹ calcium. The drop pipe, through which the extracted water is pumped to ground level, was drawn out of the well on 27 April 2015 and had deposits of pasty iron sludge on the inner surface. A sample was taken from these deposits at several points corresponding to depths between 20 and 10 m bgl. A second sample consisted of suspended iron sludge deposits that had been flushed away from the upper well intake screen and retained on a fleece filter during pumping out of the turbid water on 28 April 2015.

Enrichment and cultivation of *Ca. N. inopinata*. The biofilm used as inoculum was suspended and incubated at 46 °C with 0.5 mM NH₄Cl in a modified AOM medium⁵¹ containing (per litre): 50 mg KH₂PO₄; 75 mg KCl; 50 mg MgSO₄ × 7H₂O; 584 mg NaCl; 4 g CaCO₃ (mostly undissolved, acting as a solid buffering system and growth surface); 1 ml of specific trace element solution (TES); and 1 ml of selenium-wolfram solution (SWS)⁵². The composition of TES and SWS is described below. Both solutions were added to the autoclaved medium by sterile filtration using 0.2 µm pore-size cellulose acetate filters (Thermo Scientific). The pH of the medium was around 8.2 after autoclaving and was kept around 7.8 by the CaCO₃ buffering system during growth of the enrichment. TES contained (per litre): 34.4 mg MnSO₄ × 1H₂O; 50 mg H₃BO₃; 70 mg ZnCl₂; 72.6 mg Na₂MoO₄ × 2H₂O; 20 mg CuCl₂ × 2H₂O; 24 mg NiCl₂ × 6H₂O; 80 mg CoCl₂ × 6H₂O; 1 g FeSO₄ × 7H₂O. All salts except FeSO₄ × 7H₂O were dissolved in 997.5 ml Milli-Q water and 2.5 ml of 37% HCl was added before dissolving the FeSO₄ × 7H₂O salt. SWS contained (per litre): 0.5 g NaOH; 3 mg Na₂SeO₃ × 5H₂O; 4 mg Na₂WO₄ × 2H₂O. The primary ammonium-consuming enrichment was subsequently treated with antibiotics (one treatment with 50 mg l⁻¹ vancomycin, two treatments with 50 mg l⁻¹ bacitracin). The ammonium concentration was increased to 1 mM NH₄Cl for these and all further cultivation steps. After these treatments and repeated serial dilutions in AOM medium without antibiotics, enrichment culture ENR4 was obtained that was characterized in this study. An aliquot of ENR4 was incubated at 50 °C for four weeks and then subjected to serial dilution at 46 °C. Propagation of the most diluted (10⁻⁸) ammonia-oxidizing culture was followed by serial dilution in AOM medium containing 1 mM urea instead of ammonium. The most diluted (10⁻⁷) urea-consuming (that is, nitrifying) culture was again cultivated in AOM medium with 1 mM NH₄Cl and subjected to repeated serial dilutions, which resulted in culture ENR6 that was also characterized in this study. Enrichments ENR4 and ENR6 were further cultivated in 100 ml or 250 ml Schott bottles in AOM medium containing 1 mM NH₄Cl. To obtain enough biomass for DNA extraction, enrichment ENR4 was up-scaled in 1 l and 2 l Schott bottles. The composition of enrichment cultures was analysed by phase contrast microscopy, electron microscopy, FISH with rRNA-targeted

probes, *amoA*- and 16S rRNA-specific PCR, and metagenomics (see later for methodological details).

Physiological experiments with *Ca. N. inopinata*. To study nitrification by *Ca. N. inopinata*, an actively nitrifying ENR4 stock culture was harvested by centrifugation (9,300g, 30 min, 10 °C) and the biomass was suspended in AOM medium (see above) without ammonium. Aliquots (25 ml) of this suspension were distributed to 100 ml Schott bottles (all glassware was rinsed twice in 6 M HCl and three times in Milli-Q water, autoclaved, and dried at 60 °C before use). After addition of NH₄Cl to final concentrations of 1 mM, 0.1 mM, or 10 µM, respectively, or of NaNO₂ to a final concentration of 0.5 mM, the biomass was incubated at 46 °C for 9 h (10 µM NH₄Cl) or 48 h (other experiments) without agitation in the dark. Samples (500 µl) for chemical analyses (see below) were taken directly after ammonium or nitrite addition and during the incubations. The samples were centrifuged (22,000g, 10 min, 4 °C) to remove cells and undissolved CaCO₃ and 450 µl of the supernatant was transferred to plastic tubes and stored at -20 °C until analysis. Each incubation condition except 10 µM NH₄Cl was performed in parallel with four biological replicates (biological triplicates for 10 µM NH₄Cl), two dead biomass controls (cells were killed by autoclaving), and two abiotic controls that contained only medium and substrate, but no biomass. After the experiments, the remaining biomass was harvested by centrifugation (9,300g, 30 min, 10 °C), frozen immediately at -80 °C, and shipped on dry ice for proteome analysis. To quantify growth of *Ca. N. inopinata* by complete nitrification, culture ENR4 was incubated in mineral NOB medium, which has been used to cultivate nitrite-oxidizing *Nitrospira*²¹. In this experiment, the NOB medium was amended with ammonium instead of nitrite. The NOB medium was chosen because it contains less CaCO₃, which can affect quantitative PCR (qPCR) efficiency and accuracy. Nitrifying activity of ENR4 in NOB medium was confirmed in preceding tests. Biomass from the supernatant (without undissolved CaCO₃) from an ammonia-oxidizing culture was washed once in NOB medium, harvested by centrifugation (9,300g, 30 min, 10 °C), and prepared for incubation as described above. Following the addition of NH₄Cl to a final concentration of 0.6 mM, samples (100 µl) for quantitative PCR were taken immediately and after 4, 5, 7, and 8 days of incubation. Samples for chemical measurements (see below) were taken immediately and after 1, 4, 5, 7, and 8 days of incubation. All samples were stored at -20 °C until analysis. These incubation experiments were performed in biological triplicates. Copy numbers of the *Ca. N. inopinata amoA* gene were determined by qPCR using the newly designed *Ca. Nitrospira inopinata amoA* gene-specific primers Nino_amoA_19F (5′-ATAATCAAAGCCGCCAAGTTGC-3′) and Nino_amoA_252R (5′-AACGGCTGACGATAATTGACC-3′). The qPCR reactions were run with three technical replicates in a Bio-Rad C1000 CFX96 Real-Time PCR system, using the Bio-Rad iQ SYBR Green Supermix kit (Bio-Rad). Each qPCR reaction was performed in 20 µl reaction mix containing 10 µl SYBR Green Supermix, 2 µl of the sampled ENR4 cell suspension, 0.1 µl of each primer (50 µM), and 7.9 µl of autoclaved double-distilled ultrapure water. Cells were lysed and DNA was released for 10 min at 95 °C, followed by 43 PCR cycles of 40 s at 94 °C, 40 s at 52 °C, and 45 s at 72 °C. Plasmids carrying the *Ca. N. inopinata amoA* gene were obtained by PCR-amplifying the gene from the ENR4 culture and cloning the product into the pCR4-TOPO TA vector (Invitrogen). The M13-PCR product from these plasmids containing the *amoA* gene was used as standard for qPCR (the *amoA* copy number in the standard was calculated from DNA concentration). Tenfold serial dilutions of the standard were subjected to qPCR in triplicates to generate an external standard curve. The amplification efficiency was 92.6%, and the correlation coefficient (*r*²) of the standard curve was 0.999.

Isolation of the betaproteobacterium from ENR4. A 1 ml aliquot of the ENR4 culture was transferred to 25 ml modified AOM medium (see above) containing 6 mM sodium acetate. After three weeks of incubation at 46 °C, a 1 ml aliquot of the betaproteobacterial primary enrichment was transferred into 25 ml of fresh modified AOM medium containing 6 mM sodium acetate. After three more weeks, a 5 ml aliquot of this culture was centrifuged (9,300g, 10 min, 10 °C) and the cells were resuspended in 25 ml NOB medium (see above) containing 1 ml of SWS and 4 mM sodium acetate. Thereafter, 1 ml of the betaproteobacterial enrichment was transferred into fresh NOB medium containing 4 mM sodium acetate every 2 weeks. The fourth transfer was checked for purity by FISH with the betaproteobacterium-specific probe Nmir1009, which showed 100% overlap with the EUB338 probe mix and DAPI signals. No *Nitrospira* cells were detected by FISH in the culture.

Physiological experiments with the betaproteobacterium. To test whether the betaproteobacterium had the capability to nitrify, 20 ml of a dense pure culture of this organism was centrifuged (9,300g, 10 min, 10 °C), washed once in modified AOM medium without solid CaCO₃, and resuspended in modified AOM medium without ammonium and solid CaCO₃. Aliquots of this suspension were distributed into 100 ml Schott bottles, which had been rinsed twice in 6 M HCl, washed 3 times in Milli-Q water, closed with aluminium caps, autoclaved, and

dried at 60 °C before use. Subsequently, the following substrates were added: 1 mM NH_4Cl ; or 0.5 mM NaNO_2 and 0.1 mM NH_4Cl ; or 4 mM sodium acetate and 0.1 mM NH_4Cl (the 0.1 mM NH_4Cl was added to the nitrite and acetate incubations to provide the organism with a nitrogen source for assimilation). The biomass was incubated at 46 °C in the dark without agitation. All experiments were performed in parallel with biological triplicates. Samples (700 μl) for qPCR and chemical analyses (see below) were taken immediately after experimental set-up and after 19, 24, 30, 42, and 48 h of incubation. The samples were stored at –20 °C until analysis. Cell densities of the betaproteobacterium were quantified by qPCR targeting the *soxB* gene, which encodes the SoxB component of the periplasmic thiosulfate-oxidizing Sox enzyme complex. SoxB is a single-copy gene in the genome of the betaproteobacterium. The primers used to quantify the *soxB* gene were *soxB_F1* (5'-GGACCAGACCGCCATCACTTACCC-3') and *soxB_R1* (5'-GCACCATGTCCCCGCTTGCT-3'). The qPCR protocol and conditions were the same as described above.

Chemical analyses. Ammonium levels were measured photometrically as described previously^{53,54} with adjusted volumes of sample and reagents. Standards were prepared in AOM or NOB medium and ranged from 7.25 to 1,000 μM NH_4Cl . Nitrite concentrations were determined photometrically by the acidic Griess reaction⁵⁵. Nitrate was reduced to nitrite by vanadium chloride and measured as NO_x by the Griess assay. Nitrate concentrations were calculated from the NO_x measurements as described elsewhere⁵⁶. Standards were prepared in AOA or NOB medium and ranged from 7.8 to 1,000 μM for NO_x and from 3.9 to 500 μM for nitrite.

Replication of physiological experiments. The number of replications are detailed in the subsections for each specific experiment, and were mostly determined by the amount of biomass available for the different cultures. In all experiments, a minimum of three biological replications were used. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

FISH and microscopy. FISH with rRNA-targeted oligonucleotide probes was performed as described elsewhere⁵⁷ using the EUB338 probe mix^{58,59} for the detection of Bacteria, probes Ntspa662 and Ntspa712 specific for *Nitrospira*¹⁰, and probes Nso1225, Nso190, and NEU specific for betaproteobacterial AOB²². The betaproteobacterium in ENR4 and ENR6 was detected by FISH with the specific probe Nmir1009 (5'-CACTCCCCGCTCTCCGGG-3') with 35% of formamide in the hybridization buffer. If required, unlabelled competitor oligonucleotides were added in equimolar amounts as probes. Cells were counterstained by incubation for 5 min in a 0.1 $\mu\text{g ml}^{-1}$ DAPI (4',6-diamidino-2-phenylindole) solution. Fluorescence micrographs were recorded by using a Leica SP7 confocal laser scanning microscope equipped with a white light laser. To determine the relative abundances of *Nitrospira* and AOB in WWTP VetMed by quantitative FISH, 20 confocal images of FISH probe signals were taken at random positions in the sample and analysed as described elsewhere⁶⁰ by using the digital image analysis software daime⁶¹. For whole-cell electron microscopy, cells were positively stained with 1% (w/v) uranyl acetate. Electron microscopy of thin sections was carried out as described elsewhere⁶².

PCR assays for marker genes of AOB and AOA. To check whether the *Ca. N. inopinata* enrichments contained known AOB or AOA, PCR tests were performed using primer sets *amoA*-1F/*amoA*-2R targeting betaproteobacterial *amoA*⁶³, *CamoA*-19f/*CamoA*-616r targeting thaumarchaeal *amoA*^{33,64}, and 771F/957R for thaumarchaeal 16S rRNA genes⁶⁵ and the respective published reaction conditions. DNA was extracted for these PCR assays by using the PowerSoil DNA Isolation Kit (MoBio) according to the manufacturer's instructions.

Metaproteomic analysis. Protein extraction from concentrated ENR4 biomass, proteolytic digestion, analysis of peptide lysates by mass spectrometry (MS), processing of MS raw files, and analysis of MS spectra were carried out as described elsewhere²⁰. MS spectra were searched against a database of predicted gene products on the ENR4 metagenome scaffolds containing 12,234 sequence entries and a common Repository of Adventitious Proteins (cRAP) database using the Sequest HT algorithm. The PROPHANE pipeline (<http://www.prophane.de/index.php>) was used to classify the lowest common phylogenetic ancestor of each protein group and to calculate the normalized spectral abundance factor (NSAF).

DNA extraction for metagenomics. Biomass of enrichment ENR4 was collected from three culture bottles (samples ENR4_A, ENR4_E, ENR4_F) by centrifugation and frozen over night at –80 °C before total nucleic acids were extracted by bead beating in the presence of phosphate buffer, 10% (w/v) SDS and phenol as described elsewhere⁶⁶ (see ref. 67 for full protocol). Bead beating was repeated twice to break remaining intact cells, the supernatants from each step were pooled, and nucleic acids purified by phenol/chloroform/isoamyl alcohol and chloroform/isoamyl alcohol extraction. Nucleic acids were precipitated using 20% (w/v) polyethylene glycol, washed in ice-cold 75% (v/v) ethanol, and resuspended in

sterile 10 mM TRIS buffer. RNA was digested with RNase I (Promega) and the purity of DNA assessed by spectrophotometry. The same protocol was used to extract DNA from concentrated biomass of enrichment ENR6 (sample ENR6_N3), with the modification that bead beating was not repeated, and from an activated sludge sample of WWTP VetMed collecting only the supernatants of the second and third bead beating steps (DNA extract Vetmed_23). DNA was extracted from a second aliquot of the WWTP VetMed sample (DNA extract Vetmed_Pskit), and from pasty (sample GWW_HP_F1) or suspended (sample GWW_HP_D) iron sludge from the GWW, by using the PowerSoil DNA Isolation Kit (MoBio). DNA was extracted from all MBR samples by using the FastDNA SPIN Kit for Soil (MP Biomedicals) following the manufacturer's instructions.

Metagenome sequencing. Sequencing libraries were prepared using the Nextera or TruSeq PCR free kits (Illumina Inc.) following the manufacturer's recommendations. For the TruSeq PCR free kits, the 550 bp protocol was used with 1 μg of input DNA. The prepared libraries were sequenced using either an Illumina MiSeq with MiSeq Reagent Kit v3 (2x301 bp; Illumina Inc.) or an Illumina HiSeq2000 using the TruSeq PE Cluster Kit v3-cBot-HS and TruSeq SBS kit v3-HS sequencing kit (Illumina Inc.). Nanopore sequencing was performed in addition to facilitate completion of the *Ca. N. inopinata* genome sequence. Library preparation was done using the Nanopore Sequencing kit (SQK-MAP005, Oxford Nanopore) following the manufacturer's recommendations (v.MN005_1124_revC_02Mar2015) with shearing in an Eppendorf MiniSpin plus centrifuge at 8,000 r.p.m. and including the optional PreCR treatment step, as well as Ampure XP Bead purification after dA-tailing. The libraries were sequenced using nanopore flow cells (FLO-MAP003, Oxford Nanopore) using the MinION device (Oxford Nanopore) with the MinKNOW software (v.0.50.1.15). Flow cells were primed twice with a mixture of 3 μl Fuel Mix, 75 μl 2 × Running Buffer, and 72 μl nuclease-free water for 10 min. Libraries were prepared for loading onto the flow cell by mixing 75 μl 2 × Running Buffer, 66 μl nuclease-free water, 3 μl Fuel Mix, and 6 μl Library (Pre-sequencing Mix). A sequencing run was started (MAP_48Hr_Sequencing_Run.py) after loading the library. Additional DNA library top-ups and restart of the run script was carried out to maximize yield by allowing a new selection of active pores. Base calling was carried out using Metrichor and the 2D Basecalling workflow (Rev 1.16). Details for each metagenome can be found in Supplementary Table 1.

Metagenome scaffold assembly and binning. Paired-end Illumina reads were imported into CLC Genomics Workbench v.8.0 (CLCBio, Qiagen) and trimmed using a minimum phred score of 20 and a minimum length of 50 bp, with allowing no ambiguous nucleotides and trimming off Illumina sequencing adaptors if found. FASTQ files for the Oxford Nanopore 2D reads were obtained using the R package poRe v.0.6⁶⁸ and error corrected using Illumina reads through Proovread v.2.13⁶⁹. For each environment, all trimmed Illumina reads were co-assembled using CLCs *de novo* assembly algorithm, using a kmer of 63 and a minimum scaffold length of 1 kbp. Trimmed reads were mapped to the assembled scaffolds using CLCs map reads to reference algorithm, with a minimum similarity of 95% over 70% of the read length. Open reading frames (ORFs) were predicted in the assembled scaffolds using Prodigal⁷⁰. A set of 107 hidden Markov models (HMMs) of essential single-copy genes⁷¹ were searched against the ORFs using HMMER3 (<http://hmmer.janelia.org/>) with default settings, except option (-cut_tc) was used. Identified proteins were taxonomically classified using BLASTP against the RefSeq (v.52) protein database with a maximum *e*-value cut-off of 10^{-5} . MEGAN⁷² was used to extract class-level taxonomic assignments from the BLAST output. The script network.pl (<http://madsalbertsen.github.io/mmgenome/>) was used to extract paired-end read connections between scaffolds. PhyloPythiaS+⁷³ was used to taxonomically classify all scaffolds of selected samples. In addition, selected metagenome assemblies were binned based on ESOM maps⁷⁴. After training the ESOM using scaffolds >5 kbp and large scaffolds chopped into 5-kbp pieces, all scaffolds were projected back to the ESOM map to retrieve a single coordinate for all scaffolds. Individual genome bins were extracted using the multi-metagenome principles²³ implemented in the mmgenome R package (<http://madsalbertsen.github.io/mmgenome/>). All genome bins are fully reproducible from the raw metagenome assemblies using Rmarkdown files available on <http://madsalbertsen.github.io/mmgenome/>. The script extract.fastq.reassembly.pl was used to extract paired-end reads from the binned scaffolds, which were used for re-assembly using SPAdes⁷⁵. For selected samples, error-corrected Oxford Nanopore 2D reads were used for scaffolding using SSPACE-LongRead⁷⁶. For all genomes, quality was assessed using coverage plots through the mmgenome R package and through the use of QUAST⁷⁷ and CheckM⁷⁸. Details for each metagenome assembly can be found in Supplementary Table 2, and further details for the reconstructed bacterial genomes (including CheckM results) in Supplementary Tables 3–7. Relative genome sequence coverage was calculated as the fraction of sequence coverage of a reconstructed genome compared to the summed coverage of all genomes in these low-complexity metagenomes.

The reconstructed bacterial genomes were uploaded to the MicroScope platform⁷⁹ for automatic annotation and for manual annotation refinement¹⁷ of key pathways of *Ca. N. inopinata*.

To test for the presence of additional organisms capable of nitrification, the raw reads for each enrichment ENR4 and ENR6 were mapped to the *amoA*, *amoB*, *amoC*, *hao* and *nrxB* sequences used to generate the trees in Extended Data Figs 5b,d, 8, and 9. Reads were required to align to any one member of a target data set over at least 70% of read length with BLASTN (word size = 7). Reads that mapped with >97% nucleotide identity were automatically classified. Reads with lower identity were placed with the Evolutionary Placement Algorithm (EPA) using RAXML⁸⁰. Using this procedure, no indication was found for the presence of any nitrifier other than *Ca. N. inopinata* in these enrichments.

Sequence collection and phylogenetics. For phylogenetic analyses of AMO and HAO, full amino acid data sets were downloaded from the Pfam⁸¹ site for bacterial (pfam02461) and archaeal (pfam12942) *amoA*. Additional amino acid sequences were identified from the NCBI GenBank⁸² and the Integrated Microbial Genomes databases (IMG-ER and -MER)⁸³ that were returned using the search words 'ammonia, methane, amo, pmo or monooxygenase' (GenBank) or had been annotated with one of the target pfams (IMG). A BLASTP⁸⁴ search was performed using the *Ca. Nitrospira inopinata amoA* sequence as a query, word size = 2, BLOSUM 45, E = 10 and the top 1,000 returned sequences were downloaded. Comparable procedures were performed to generate a comprehensive set of *amoB* (pfam04744) and *amoC* (pfam04896) sequences. For construction of the *hao* (pfam13447) data set, query words were changed to 'hydroxylamine' and 'Hao'. For each gene set, amino acid sequences were filtered using hmmersearch (<http://hmmer.janelia.org/>) with the respective pfam HMMs, requiring an expect value < 0.001. Amino acid sequences were clustered at 75% identity using USEARCH⁸⁵ and aligned using Mafft⁸⁶. Phylogenetic trees were calculated using PhyloBayes⁸⁷, running 5 independent chains for 21,000 cycles each, using 11,000 cycles for burn-in and sampling every 20 cycles. Sequences that mapped to centroids that clustered within the comammox clade were used for additional phylogenetic calculations along with an outgroup of 27 betaproteobacterial *amoA* and 29 diverse *pmoA* sequences. Corresponding nucleotide sequences for this set were aligned according to their amino acid translations using MUSCLE⁸⁸ and manually corrected for frameshifts. Nucleotide alignments were then used for constructing consensus trees in PhyloBayes, running 5 independent chains for 21,000 cycles each, using 11,000 cycles for burn-in and sampling every 20 cycles.

To estimate relative abundances of *amoA* genes, comammox-type *amoA* sequences were identified from three publicly available Rifle soil metagenomic data sets (3300002121, 3300002122 and 3300002124) available within IMG. Functional profiles were generated within IMG using pfam12942 (archaeal *amoA*) and pfam02461 (bacterial *amoA/pmoA*) against the assembly and unassembled reads. All identified *amoA/pmoA* nucleotide sequences were downloaded as nucleic acid sequences and added to the existing *amoA* alignment used to generate Extended Data Fig. 8 with the -add option in Mafft. EPA in RAXML was used to assign downloaded sequences into the reference tree that is the basis for Extended Data Fig. 8. *AmoA* abundance for each *amoA* type (comammox, archaeal, betaproteobacterial AOB) was estimated by taking the sum of the estimated copy numbers of each assembled *amoA* gene of a given type as well as the number of unassembled reads assigned to a given *amoA* type.

Comammox, betaproteobacterial, and archaeal *amoA* sequences from the metagenomes of WWTP VetMed and the GWW were identified using the same procedure as above. Comammox *amoA* read abundances were then used to calculate an estimate of the fraction of *Nitrospira* that are comammox. *AmoA* was assumed to be a single copy gene in all comammox (as it is in *Ca. N. inopinata*). Total *Nitrospira* were enumerated by mapping raw reads from metagenomic samples using the first 700 nucleotides of the predicted ATP-citrate lyase subunit beta (*aclB*) gene from *Ca. N. inopinata*. Reads were required to align to *Ca. N. inopinata aclB* over at least 70% of read length and with >60% alignment identity with BLASTN (word size = 7). *aclB* was chosen on the basis that this gene has a restricted taxonomic distribution, encodes a key enzyme of the reductive tricarboxylic acid cycle employed by all known *Nitrospira* for CO₂ fixation, and is present in single copy within known *Nitrospira* genomes. To test its utility, all 150 nt segments (pos 1:150, 2:151...1,051:1,200) of the *Ca. N. inopinata aclB* gene was used as a query against the nr database (BLAST, word size = 7, 70% read length and 60% alignment identity). Over the first 700 nucleotides of the *aclB* gene, test fragments mapped only to reference *Nitrospira* organisms. Downstream of this region, the *aclB* mapping was less specific, mapping to *Nitrospira* and *Chlorobi* with high (>90%) identity. Coverage of each gene was calculated by dividing the number of mapped reads by gene length of the query (843 nt for comammox *amoA* and 700 nt for *Nitrospira aclB*). Adjusted coverage was calculated by dividing gene coverage by total number of reads in the metagenomic data set. Ratios discussed in the main text are then the adjusted coverage of comammox (as calculated from

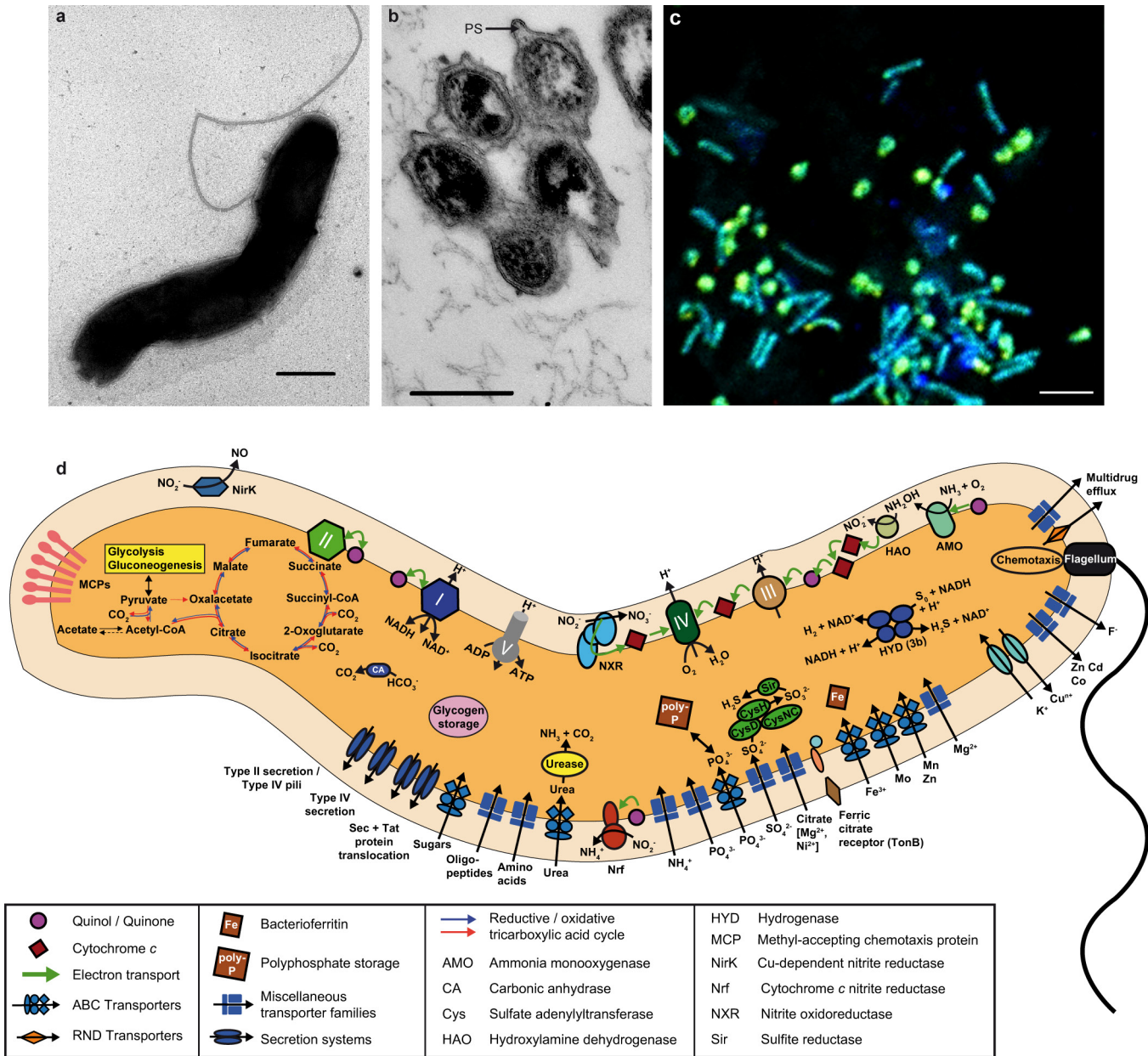
comammox *amoA*) divided by the adjusted coverage for all *Nitrospira* (as calculated from *aclB*).

For phylogenetic analyses of NXR, the *NxrA* and *nrxB* sequences of *Ca. N. inopinata* were imported into existing *NxrA*¹⁷ and *nrxB*⁸ sequence databases using the software ARB⁸⁹. *NxrA* sequences were aligned using Mafft, *nrxB* sequences were manually aligned according to their amino acid translations. Maximum likelihood trees were calculated using RAXML with the GAMMA model of rate heterogeneity using empirical base frequencies and the LG substitution model (*NxrA*) or with the GAMMA model of rate heterogeneity and the GTR substitution model (*nrxB*). Bayesian inference trees were calculated using PhyloBayes, running 3 independent chains for 32,200 cycles each, using 6,440 cycles for burn-in (*NxrA*) or 3 independent chains for 35,500 cycles each, using 7,000 cycles for burn-in (*nrxB*). *Nitrospira* 16S rRNA genes from this study were added to an existing *Nitrospira* 16S rRNA sequence database and aligned in ARB. Phylogenetic trees were calculated using RAXML with the GAMMA model of rate heterogeneity and the GTR substitution model, and using MrBayes⁹⁰ v.3.2.1, running 4 independent chains for 5 million generations each, with 1.25 million cycles for burn-in and sampling every 100 generations.

Average nucleotide identity and tetranucleotide signature analyses. Pairwise average nucleotide identity (ANI) values were calculated for comammox *Nitrospira* genomes using BLAST (ANi) in JSpecies⁹¹. Genome-wide tetranucleotide signatures were calculated for the forward and reverse strand for each genome with the oligonucleotideFrequency (width = 4) command using the Biostrings package in R⁹². Tetranucleotide patterns were also calculated across the length of the genome with a sliding window of 5 kb (step = 1 kb). The tetranucleotide pattern for each window was compared to the global tetranucleotide signature by calculating the Pearson correlation (*r*) of log(1+counts) of each window against the log(1+counts) of the global signature. *P* values, indicating a significantly low correlation for tetranucleotide signature of a window, were calculated by modelling 1 - *r* across all windows as a log-normal distribution. Multiple testing was accounted for by using the Benjamini-Hochberg procedure with a false discovery rate of 5%.

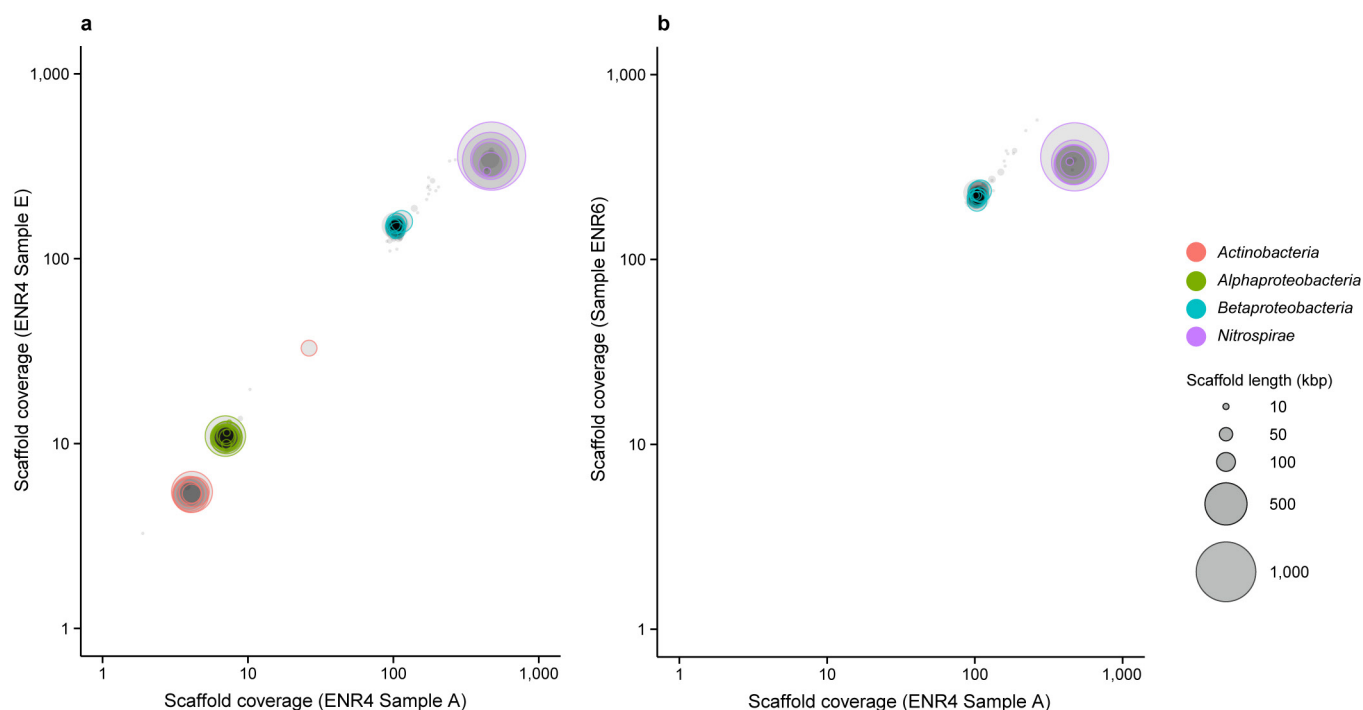
51. Lebedeva, E. V. et al. Enrichment and genome sequence of the group I.1a ammonia-oxidizing Archaeon "Ca. Nitrosotenuis uzonensis" representing a clade globally distributed in thermal habitats. *PLoS One* **8**, e80835 (2013).
52. Widdel, F. *Anaerobier Abbau von Fettsäuren und Benzoesäure durch neu isolierte Arten Sulfat-reduzierender Bakterien*. PhD thesis, Univ. Göttingen (1980).
53. Kandel, E. & Gerber, H. Short-term assay of soil urease activity using colorimetric determination of ammonium. *Biol. Fertil. Soils* **6**, 68–72 (1988).
54. Hood-Nowotny, R., Hinko-Najera Umana, N., Inselbacher, E., Oswald-Lachouani, P. & Wanek, W. Alternative methods for measuring inorganic, organic, and total dissolved nitrogen in soil. *Soil Sci. Soc. Am. J.* **74**, 1018–1027 (2010).
55. Griess-Romijn van Eck, E. Physiological and chemical tests for drinking water. NEN 1056 IV-2 Nederlands Normalisatie Instituut, Rijswijk, The Netherlands (1966).
56. Miranda, K. M., Espey, M. G. & Wink, D. A. A rapid, simple spectrophotometric method for simultaneous detection of nitrate and nitrite. *Nitric Oxide* **5**, 62–71 (2001).
57. Daims, H., Stoecker, K. & Wagner, M. in *Molecular Microbial Ecology* (eds Osborn, A. M. & Smith, C. J.) 213–239 (Bios-Garland, 2005).
58. Daims, H., Brühl, A., Amann, R., Schleifer, K.-H. & Wagner, M. The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. *Syst. Appl. Microbiol.* **22**, 434–444 (1999).
59. Amann, R. I. et al. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl. Environ. Microbiol.* **56**, 1919–1925 (1990).
60. Daims, H. Use of fluorescence *in situ* hybridization and the *daime* image analysis program for the cultivation-independent quantification of microorganisms in environmental and medical samples. *Cold Spring Harb. Protoc.* **2009**, t5253 (2009).
61. Daims, H., Lückner, S. & Wagner, M. *daime*, a novel image analysis program for microbial ecology and biofilm research. *Environ. Microbiol.* **8**, 200–213 (2006).
62. Sorokin, D. Y. et al. Nitrification expanded: discovery, physiology and genomics of a nitrite-oxidizing bacterium from the phylum *Chloroflexi*. *ISME J.* **6**, 2245–2256 (2012).
63. Rothauwe, J.-H., Witzel, K.-P. & Liesack, W. The ammonia monooxygenase structural gene *amoA* as a functional marker: molecular fine-scale analysis of natural ammonia-oxidizing populations. *Appl. Environ. Microbiol.* **63**, 4704–4712 (1997).
64. Tournai, M., Freitag, T. E., Nicol, G. W. & Prosser, J. I. Growth, activity and temperature responses of ammonia-oxidizing archaea and bacteria in soil microcosms. *Environ. Microbiol.* **10**, 1357–1364 (2008).
65. Ochsenreiter, T., Selez, D., Quaiser, A., Bonch-Osmolovskaya, L. & Schleper, C. Diversity and abundance of *Crenarchaeota* in terrestrial habitats studied by 16S RNA surveys and real time PCR. *Environ. Microbiol.* **5**, 787–797 (2003).

66. Angel, R., Claus, P. & Conrad, R. Methanogenic archaea are globally ubiquitous in aerated soils and become active under wet anoxic conditions. *ISME J.* **6**, 847–862 (2012).
67. Angel, R. Total nucleic acid extraction from soil. *Protoc. Exch.* <http://dx.doi.org/10.1038/protex.2012.046> (2012).
68. Watson, M. *et al.* poRe: an R package for the visualization and analysis of nanopore sequencing data. *Bioinformatics* **31**, 114–115 (2015).
69. Hackl, T., Hedrich, R., Schultz, J. & Förster, F. proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. *Bioinformatics* **30**, 3004–3011 (2014).
70. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
71. Dupont, C. L. *et al.* Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J.* **6**, 1186–1199 (2012).
72. Huson, D. H., Mitra, S., Ruscheweyh, H. J., Weber, N. & Schuster, S. C. Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* **21**, 1552–1560 (2011).
73. Gregor, I., Dröge, J., Schirmer, M., Quince, C. & McHardy, A. C. PhyloPythiaS+: A self-training method for the rapid reconstruction of low-ranking taxonomic bins from metagenomes. Preprint at <http://arxiv.org/abs/1406.7123> (2014).
74. Dick, G. J. *et al.* Community-wide analysis of microbial genome sequence signatures. *Genome Biol.* **10**, R85 (2009).
75. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
76. Boetzer, M. & Pirovano, W. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics* **15**, 211 (2014).
77. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
78. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
79. Vallenet, D. *et al.* MicroScope—an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucleic Acids Res.* **41**, D636–D647 (2013).
80. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
81. Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222–D230 (2014).
82. Benson, D. A. *et al.* GenBank. *Nucleic Acids Res.* **43**, D30–D35 (2015).
83. Markowitz, V. M. *et al.* IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res.* **42**, D560–D567 (2014).
84. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
85. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
86. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
87. Lartillot, N., Lepage, T. & Blanquart, S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288 (2009).
88. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
89. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (2004).
90. Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
91. Richter, M. & Rosselló-Móra, R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl Acad. Sci. USA* **106**, 19126–19131 (2009).
92. The R Project. R: A language and environment for statistical computing. <https://www.r-project.org/> (R Foundation for Statistical Computing, 2013).



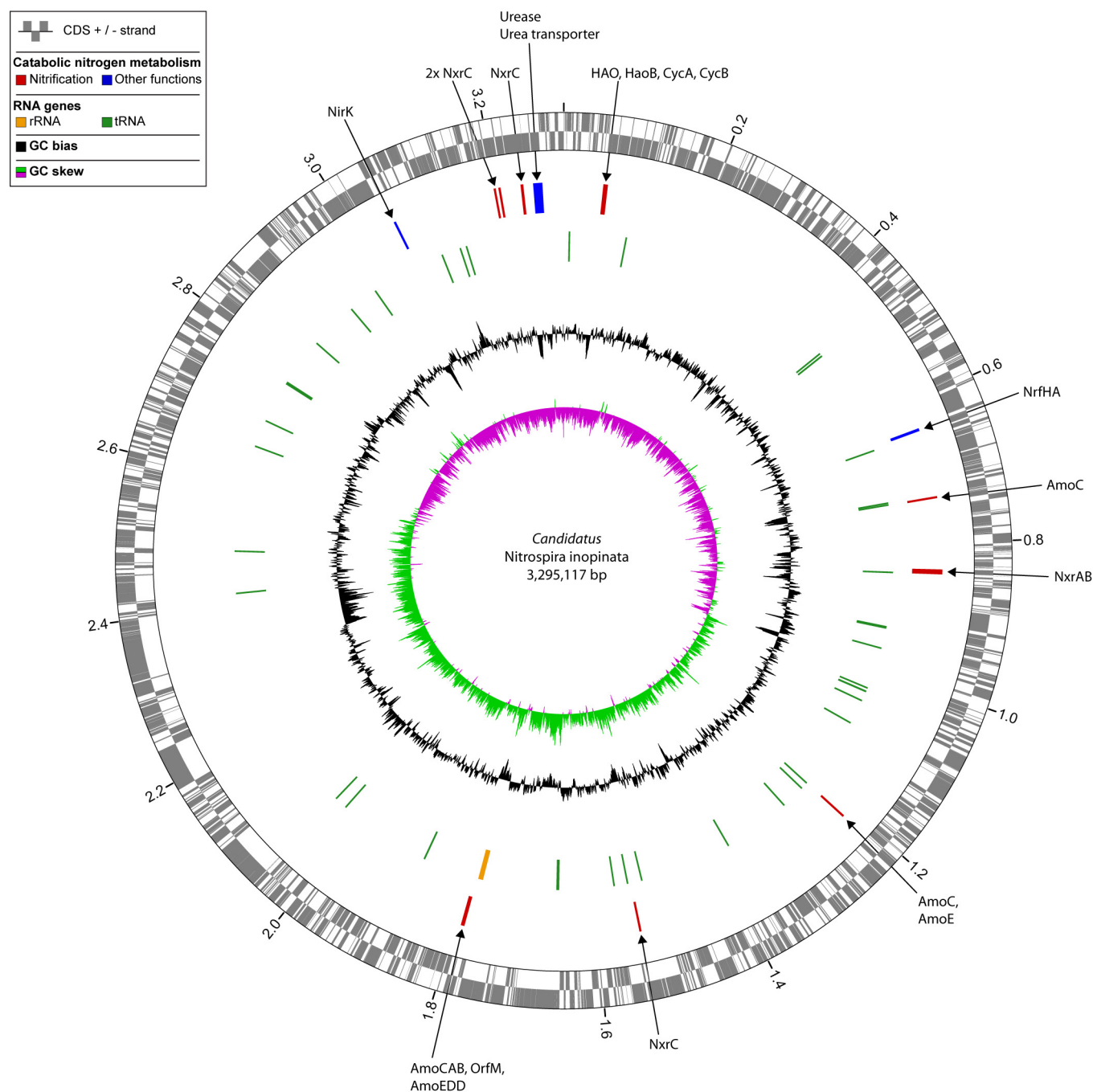
Extended Data Figure 1 | Photomicrographs and cell diagram of *Ca. Nitrospira inopinata*. **a**, Transmission electron micrograph of a spiral-shaped cell with a flagellum. The size of *Ca. N. inopinata* cells is 0.18 to 0.3 μm in width and 0.7 to 1.6 μm in length. Scale bar represents 200 nm. **b**, Transmission electron micrograph of a thin section preparation. Microcolony showing the wide periplasmic space (PS), which is a characteristic feature of *Nitrospira*¹⁵. Scale bar represents 200 nm. **c**, Fluorescence image of cells from enrichment ENR4 after hybridization with oligonucleotide probes targeting *Nitrospira* (Ntspa662 and Ntspa712

both labelled with Cy3, red), the betaproteobacterium (Nmir1009 labelled with Cy5, blue), and *Bacteria* (EUB338 probe mix labelled with FLUOS, green). *Ca. N. inopinata* cells and microcolonies appear yellow and the betaproteobacterial cells appear cyan due to simultaneous hybridization to the respective specific probe and the EUB338 probe mix. Scale bar represents 2 μm . **d.** Cell metabolic cartoon constructed from the annotation of the *Ca. N. inopinata* genome. Enzyme complexes of the electron transport chain are labelled by Roman numerals.



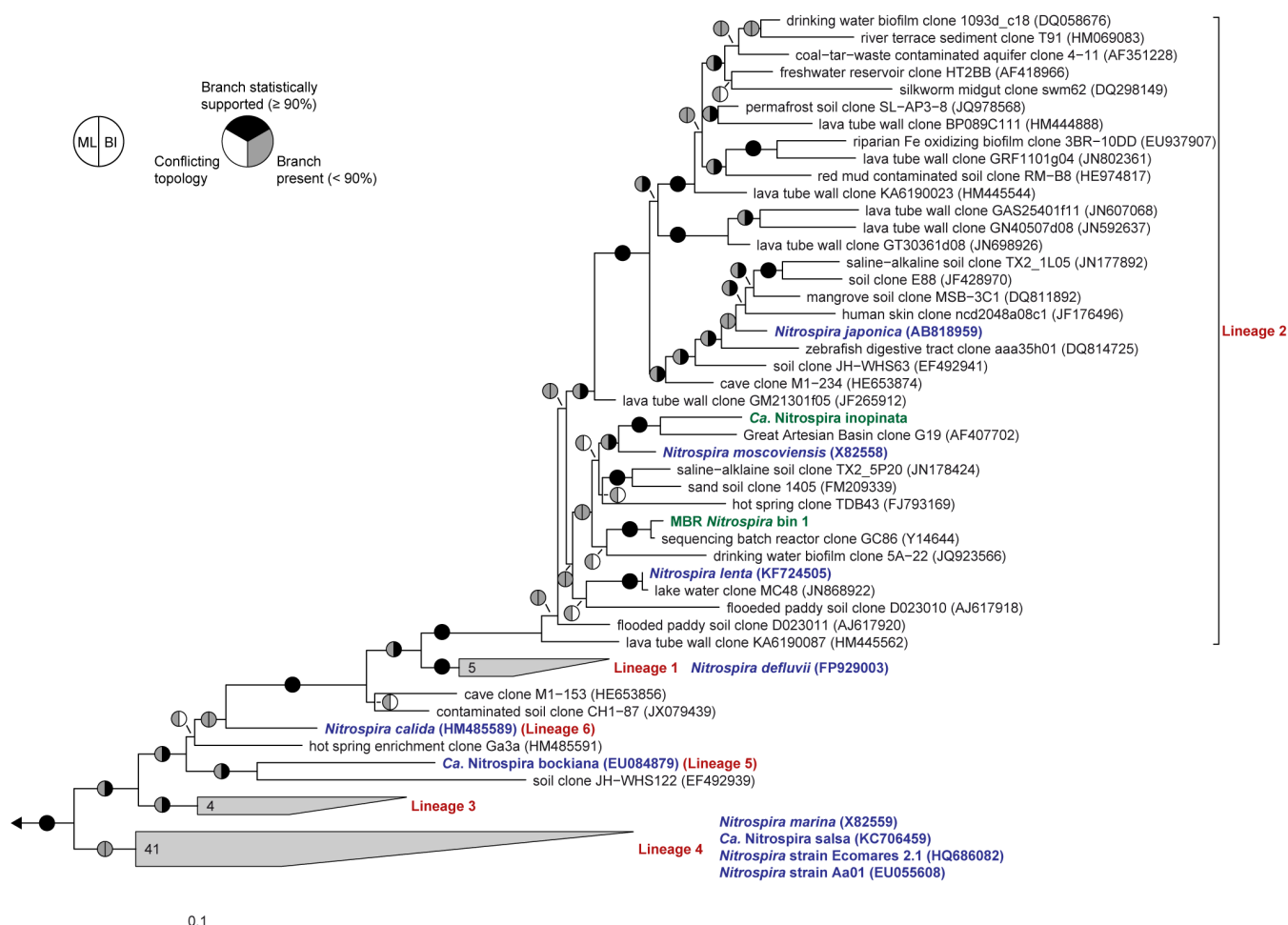
Extended Data Figure 2 | Sequence composition-independent binning of the metagenome scaffolds from the nitrifying enrichment cultures. Circles represent scaffolds, scaled by the square root of their length. Only scaffolds ≥ 5 kbp are shown. Clusters of similarly coloured circles represent potential genome bins. These differential coverage plots were the starting points for further refinement and finishing of genome assemblies as described elsewhere²³. **a**, Binning of the scaffolds from enrichment

culture ENR4 containing *Ca. N. inopinata* and three heterotrophic populations related to the Betaproteobacteria, Alphaproteobacteria, and Actinobacteria. **b**, Binning of the scaffolds from enrichment culture ENR6 containing only *Ca. N. inopinata* and the betaproteobacterial accompanying heterotrophic organism. Enrichment ENR4, sample A, was used for comparison in differential coverage binning of culture ENR6.



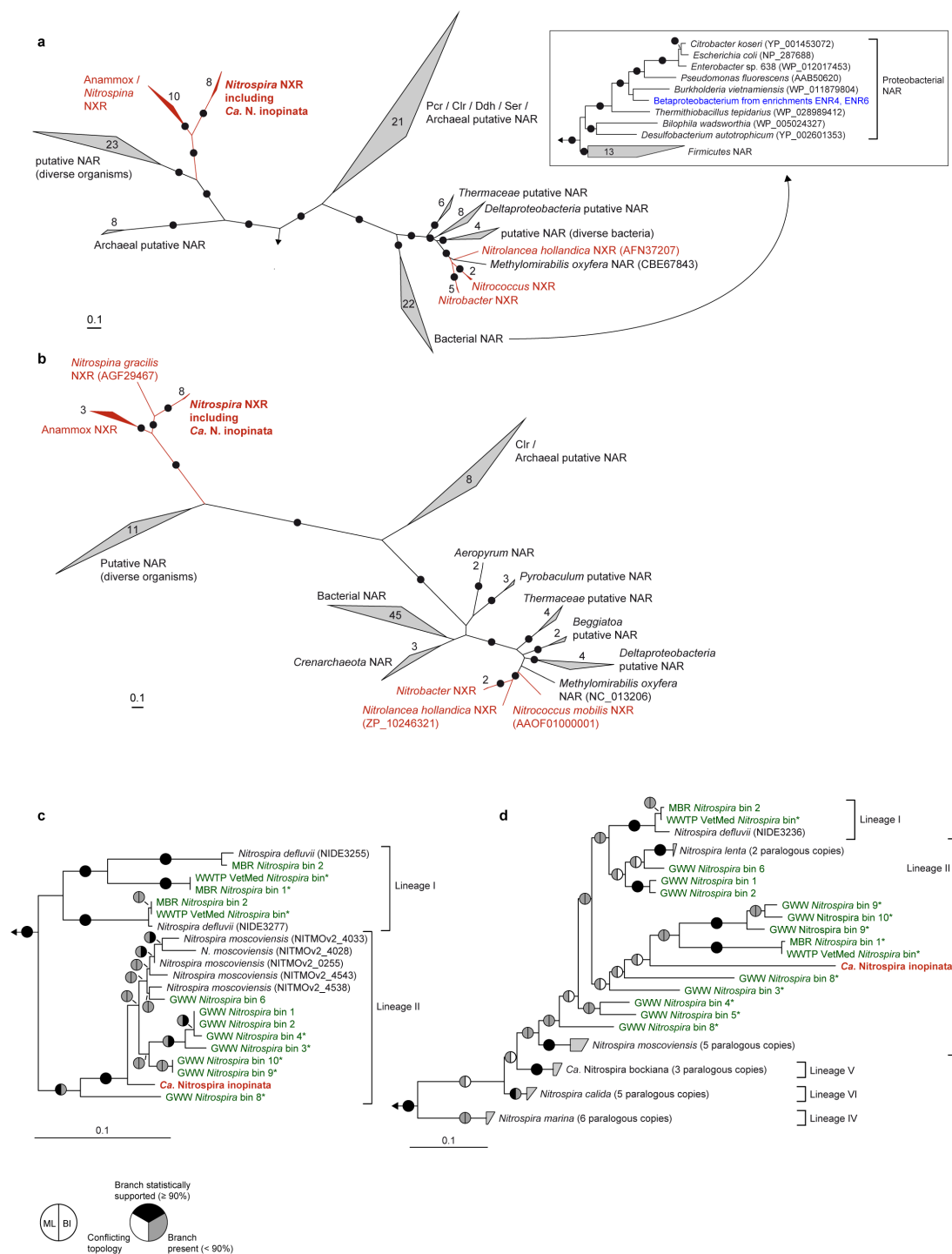
Extended Data Figure 3 | Circular representation of the *Ca. N. inopinata* chromosome. Predicted coding sequences (CDS; rings 1+2), genes of enzymes involved in nitrification and other pathways of catabolic nitrogen metabolism (ring 3), RNA genes (ring 4), and local nucleotide composition measures (rings 5+6) are shown. Very short features were enlarged to enhance visibility. Clustered genes, such as several transfer

RNA genes, may appear as one line owing to space limitations. The tick interval is 0.2 Mbp. Amo, ammonia monooxygenase; HAO, hydroxylamine dehydrogenase; CycA and CycB, tetraheme *c*-type cytochromes that form the hydroxylamine ubiquinone redox module together with HAO; NirK, Cu-dependent nitrite reductase; Nrf, cytochrome *c* nitrite reductase; Nxr, nitrite oxidoreductase; Orf, open reading frame.



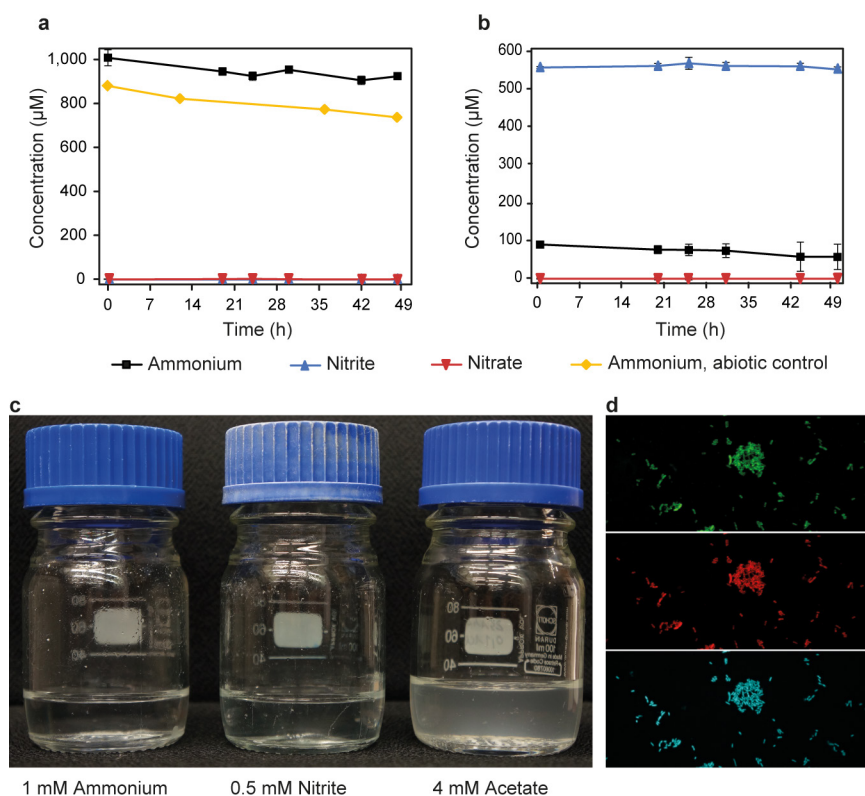
Extended Data Figure 4 | Phylogenetic affiliation of *Ca. N. inopinata*. The maximum likelihood tree, which is based on 16S ribosomal RNA sequences of cultured and uncultured representative members of the genus *Nitrospira*, shows that the comammox organism *Ca. N. inopinata* (highlighted green) is a member of *Nitrospira* lineage II. Another 16S rRNA gene sequence was extracted from MBR metagenomic *Nitrospira* bin 1 (also highlighted green). This sequence bin also contained *amo* and *hao* genes (main text Fig. 1, Extended Data Figs 8 and 9). The cultured

Nitrospira strains other than *Ca. N. inopinata*, which are not known to use ammonia as a source of energy and reductant, are highlighted blue. *Nitrospira* lineages are labelled red. Pie charts indicate statistical support of branches based on maximum likelihood (ML; 1,000 bootstrap iterations) and Bayesian inference (BI; posterior probability, 4 independent chains). In total, 95 taxa and 1,543 nucleotide sequence alignment positions were considered. Numbers in wedges indicate the numbers of taxa. The scale bar indicates 0.1 estimated substitutions per nucleotide.



Extended Data Figure 5 | Phylogeny of NXR from *Ca. N. inopinata* and related proteins. **a, b**, Maximum likelihood trees showing the alpha (**a**) and beta (**b**) subunits of selected enzymes from the DMSO reductase type II family. Names of validated enzymes are indicated (Clr, chlorate reductase; Ddh, dimethylsulfide dehydrogenase; NAR, nitrate reductase; NXR, nitrite oxidoreductase; Pcr, perchlorate reductase; Ser, selenate reductase). More distantly related molybdoenzymes were used as outgroup. Black dots on branches indicate high maximum likelihood bootstrap support ($\geq 90\%$; 1,000 iterations). Known NXR forms are highlighted in red. The inset in **a** contains a subtree, which shows the phylogenetic affiliation of the NAR of the betaproteobacterium from enrichments ENR4 and ENR6 (highlighted in blue) with canonical nitrate reductases of Proteobacteria. In total, 1,279 (**a**) and 556 (**b**) amino acid sequence alignment positions, and 134 (**a**) and 99 (**b**) taxa (including outgroups), were considered. **c, d**, Maximum likelihood trees showing only *Nitrospira* NxrA (**c**) and *nxB* (**d**) phylogenies. The tree in **d** was

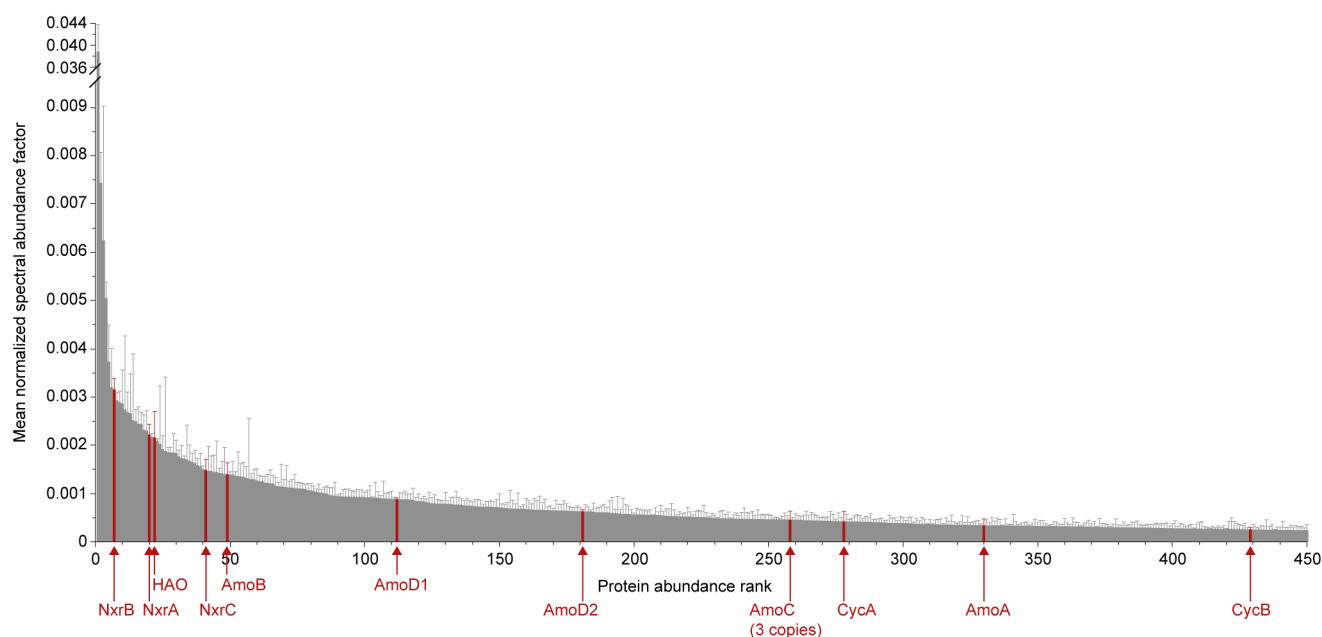
calculated using nucleotide sequences aligned according to their amino acid translations. *Ca. N. inopinata* is highlighted in red, sequences from metagenomic *Nitrospira* bins obtained in this study are highlighted in green. Asterisks mark metagenomic bins that also contain *amo* genes. Metagenomic bins are numbered as in Supplementary Table 8. Sublineages of the genus *Nitrospira* are indicated. As recognized earlier⁸, lineage II is paraphyletic with respect to lineage I in *nxB* phylogenies, but differentiation of the lineages is stable. Pie charts indicate statistical support of branches based on maximum likelihood (ML; 1,000 bootstrap iterations) and Bayesian inference (BI; posterior probability, 3 independent chains). In total, 1,279 amino acid sequence alignment positions (**c**) and 1,290 nucleotide sequence alignment positions (**d**), and 30 (**c**) and 40 (**d**) taxa (including outgroups), were considered. All panels: numbers in or next to wedges indicate the numbers of taxa. The scale bars indicate 0.1 estimated substitutions per residue.



Extended Data Figure 6 | Absence of nitrifying activity in the betaproteobacterium found in enrichments ENR4 and ENR6.

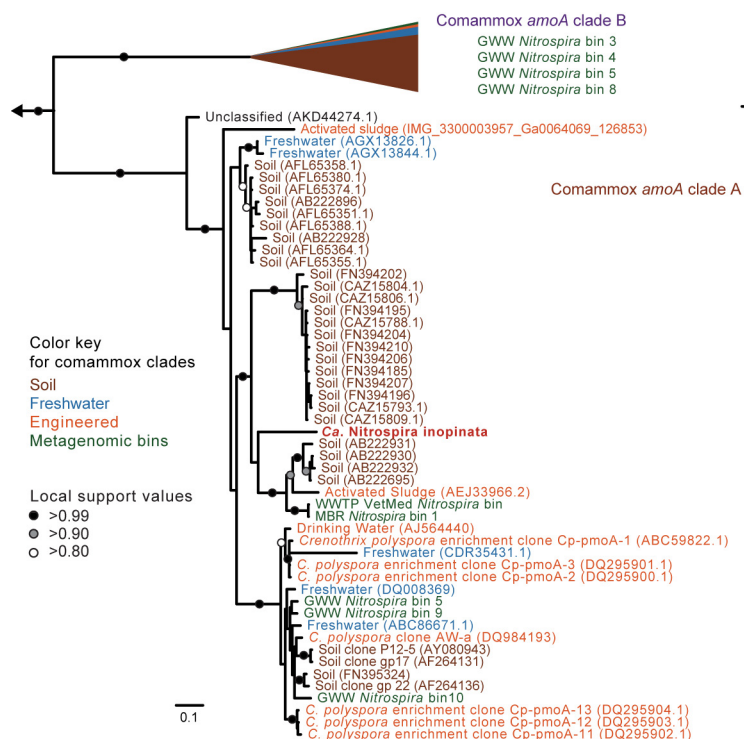
a, b, Incubation of a pure culture of the betaproteobacterium in mineral medium containing 1 mM ammonium (**a**) or 0.5 mM nitrite plus 0.1 mM ammonium as nitrogen source (**b**). No conversion of ammonium to nitrite or nitrate, or of nitrite to nitrate, was observed. Data points in **a** and **b** show means, error bars show 1 s.d. of $n = 3$ biological replicates. If not visible, error bars are smaller than symbols. The mean initial densities of the cultures, as determined by qPCR of the single-copy *soxB* gene, were 7.15 ± 0.01 ($\log(\text{soxB copies ml}^{-1})$, 1 s.d., $n = 3$) for the 1 mM ammonium experiment (**a**) and 7.22 ± 0.02 ($\log(\text{soxB copies ml}^{-1})$, 1 s.d., $n = 3$) for the 0.5 mM nitrite plus 0.1 mM ammonium experiment (**b**). After 48 h of incubation, the mean densities were 7.06 ± 0.10 and 7.15 ± 0.29 , respectively. A slight decrease in the ammonium concentration was observed in these experiments and also in an abiotic control incubation containing only medium and 1 mM ammonium, but no cells (data points for this control show means of two technical replicates). It might be explained by adsorption of ammonium to the glass bottles or by outgassing

of NH_3 . **c**, Photographs of incubation bottles after 53 h of incubation. The mean optical density at 600 nm (OD_{600}) of the cultures at this time point was 0.006 ± 0.003 (1 s.d., $n = 3$) for the 1 mM ammonium experiment and 0.007 ± 0.008 (1 s.d., $n = 3$) for the 0.5 mM nitrite plus 0.1 mM ammonium experiment. Control incubations were carried out in medium containing 4 mM acetate and 0.1 mM ammonium as nitrogen source for assimilation (three biological replicates). The inoculum for these cultures was 2.5-fold diluted compared to the experiments with ammonium or nitrite. After incubation, the acetate-grown cultures were visibly turbid with a mean OD_{600} of 0.068 ± 0.011 (1 s.d., $n = 3$) and the mean density was 8.12 ± 0.03 ($\log(\text{soxB copies ml}^{-1})$, 1 s.d., $n = 3$). Thus, the culture of the betaproteobacterium, which was used to inoculate all experiments, was physiologically active and grew on acetate. **d**, Fluorescence images showing the culture of the betaproteobacterium after FISH with the EUB338 probe mix (labelled with FLUOS, green), probe Nmir1009 that is specific for this organism (labelled with Cy3, red), and DAPI counterstaining (blue). The images show the same field of view after splitting the colour channels. According to FISH, all detected cells were the betaproteobacterium.



Extended Data Figure 7 | Protein abundance levels of *Ca. N. inopinata* during growth on ammonium. Displayed are the 450 most abundant proteins from *Ca. N. inopinata* in the metaproteome from culture ENR4 after incubation with 1 mM ammonium for 48 h. Red arrows and labels highlight key proteins for ammonia and nitrite oxidation. Columns show the mean normalized spectral abundance factor (NSAF), error bars

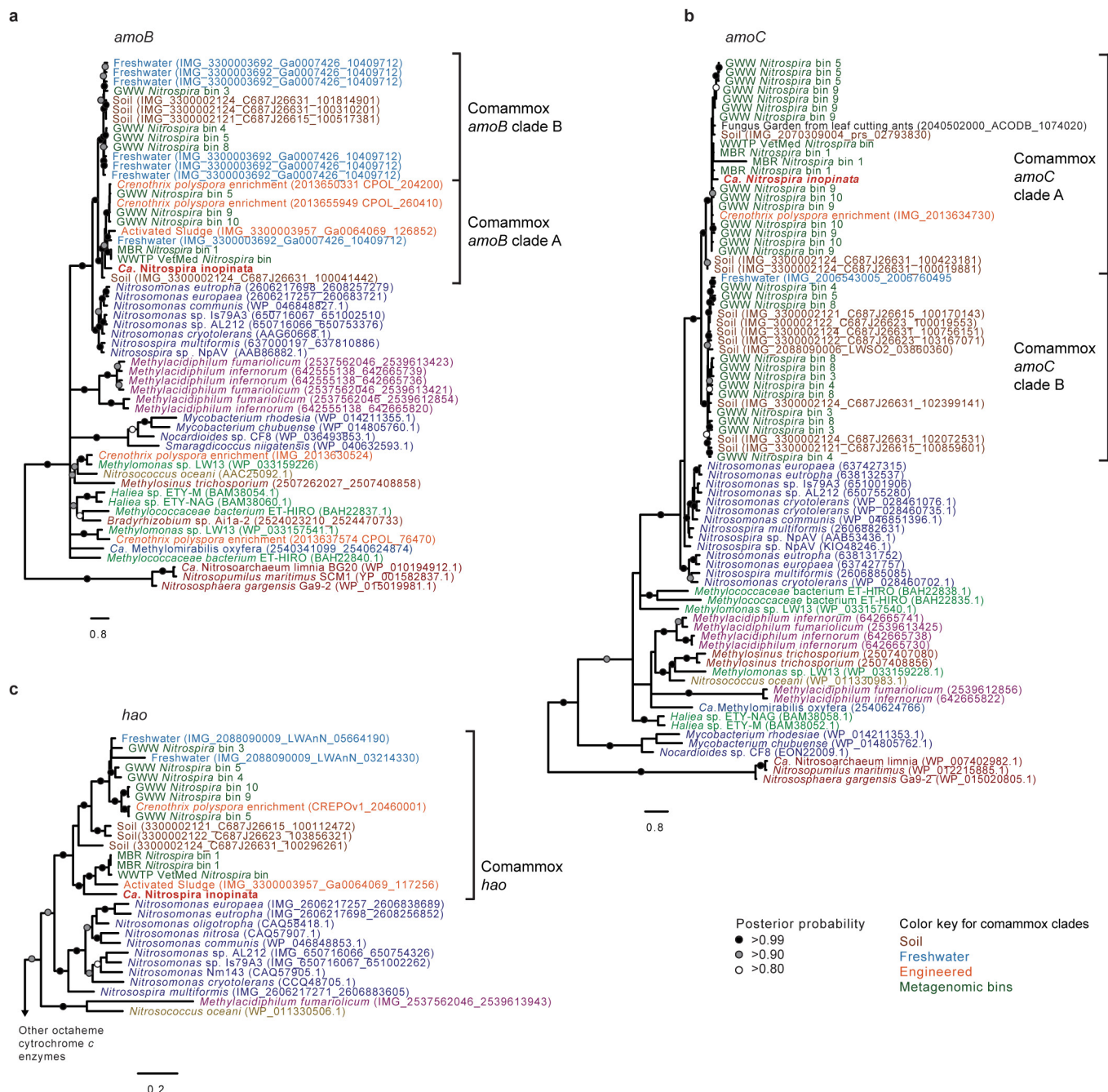
show 1 s.d. of $n = 4$ biological replicates. In total 1,083 proteins in the metaproteome were unambiguously assigned to *Ca. N. inopinata*. Only one of the four putative NXR gamma subunits (NxrC) was among the top 450 expressed proteins. The other three NxrC candidates ranked at positions 561, 605 and 931. The AmoE1 protein was ranked at position 520, and HaoB at position 653.



Extended Data Figure 8 | Phylogenetic affiliation of comammox *amoA* sequences to *amoA* sequences from different environments.

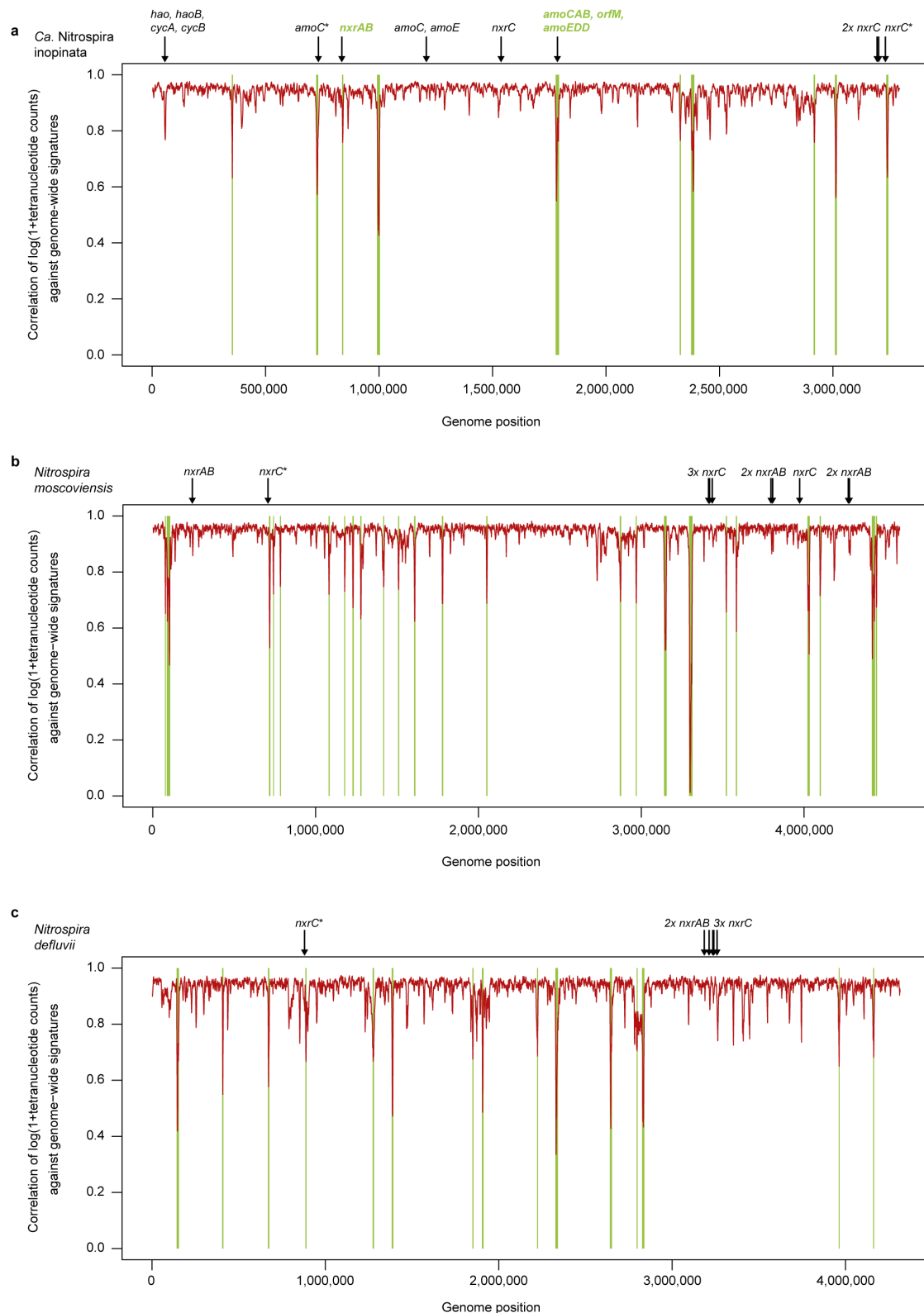
Bayesian inference tree showing the phylogenetic relationship of the *amoA* sequences from *Ca. N. inopinata* and metagenomic bins from this study (224 taxa, 939 nucleotide alignment positions). *Ca. N. inopinata* clusters confidently into comammox *amoA* clade A. Comammox *amoA* clade B (116 taxa) has been collapsed for clarity and the proportion of database

sequences from soil (95 taxa), freshwater (13 taxa), and engineered environments (4 taxa) is represented as a proportion of the collapsed clade. *AmoA* from the metagenomic *Nitrospira* bins generated for this study (5 taxa in clade A, 4 taxa in clade B) are numbered as in Supplementary Table 8. Scale bar indicates estimated change per nucleotide. The outgroup consists of 27 betaproteobacterial *amoA* and 29 diverse *pmoA* sequences.



Extended Data Figure 9 | Phylogenetic relationship of comammox *amoB*, *amoC* and *hao* sequences to corresponding gene family members. Trees were calculated with PhyloBayes using nucleotide sequences aligned according to their amino acid translations. Support values indicate the consensus probability from 5 independent chains. Sequences outside the comammox clades are coloured as in main text Fig. 3. Metagenomic bins are numbered as in Supplementary Table 8.

Scale bars indicate the estimated substitutions per nucleotide. **a**, Phylogenetic relationship of *Ca. N. inopinata amoB* to other *amoB* and *pmoB* genes (57 taxa, 1,518 alignment positions). **b**, Phylogenetic relationship of *Ca. N. inopinata amoC* to other *amoC* and *pmoC* genes (81 taxa, 993 alignment positions). **c**, Phylogenetic relationship of *Ca. N. inopinata hydroxylamine dehydrogenase (hao)* to other *hao* genes (37 taxa, 2,875 alignment positions).



Extended Data Figure 10 | Genome-wide tetranucleotide analysis of *Ca. N. inopinata* and other *Nitrospira*. Correlation of tetranucleotide patterns in a 5 kb sliding window (step size 1 kb) against genome-wide tetranucleotide signatures. The positions of key nitrification genes are indicated. Regions where the tetranucleotide patterns significantly deviate from the genome-wide signature, and nitrification genes located in such regions, are highlighted in green. Asterisks mark genes that are outside significantly deviating regions but may appear to be inside due to space limitations in the figure. **a**, *Ca. N. inopinata* (member of *Nitrospira* lineage II). The *hao*, *cycA*, and *cycB* genes are

located in a region whose tetranucleotide pattern deviates slightly but not significantly from the genome-wide signature. The *P* value cut-off from the Benjamini–Hochberg procedure, indicating a significantly low correlation for a window’s tetranucleotide signature, was 0.00065 for this genome. **b**, *N. moscoviensis* (member of *Nitrospira* lineage II). The *P* value cut-off for this genome was 0.0013. **c**, *N. defluvii* (member of *Nitrospira* lineage I). The *P* value cut-off for this genome was 0.00072. In *N. moscoviensis* (**b**) and *N. defluvii* (**c**), all *nxr* genes are outside regions with significantly deviating tetranucleotide patterns.

Δ F508 CFTR interactome remodelling promotes rescue of cystic fibrosis

Sandra Pankow^{1*}, Casimir Bamberger^{1*}, Diego Calzolari¹, Salvador Martínez-Bartolomé¹, Mathieu Lavallée-Adam¹, William E. Balch² & John R. Yates III¹

Deletion of phenylalanine 508 of the cystic fibrosis transmembrane conductance regulator (Δ F508 CFTR) is the major cause of cystic fibrosis, one of the most common inherited childhood diseases. The mutated CFTR anion channel is not fully glycosylated and shows minimal activity in bronchial epithelial cells of patients with cystic fibrosis. Low temperature or inhibition of histone deacetylases can partly rescue Δ F508 CFTR cellular processing defects and function. A favourable change of Δ F508 CFTR protein–protein interactions was proposed as a mechanism of rescue; however, CFTR interactome dynamics during temperature shift and inhibition of histone deacetylases are unknown. Here we report the first comprehensive analysis of the CFTR and Δ F508 CFTR interactome and its dynamics during temperature shift and inhibition of histone deacetylases. By using a novel deep proteomic analysis method, we identify 638 individual high-confidence CFTR interactors and discover a Δ F508 deletion-specific interactome, which is extensively remodelled upon rescue. Detailed analysis of the interactome remodelling identifies key novel interactors, whose loss promote Δ F508 CFTR channel function in primary cystic fibrosis epithelia or which are critical for CFTR biogenesis. Our results demonstrate that global remodelling of Δ F508 CFTR interactions is crucial for rescue, and provide comprehensive insight into the molecular disease mechanisms of cystic fibrosis caused by deletion of F508.

Cystic fibrosis (CF) is one of the most common inherited childhood diseases, with about 10 million carriers in the USA alone. The disease is caused by mutation of the *CFTR* gene, which encodes an ion channel critical for salt homeostasis of several polarized epithelial tissues including the lung, intestine, pancreas and kidney. Disturbed salt homeostasis in patients with CF leads to impaired clearance of mucus from the respiratory tract, subsequent chronic lung infections and inflammation, and eventual respiratory failure^{1,2}. The most prevalent mutation, occurring in more than 70% of patients, is an in-frame-deletion of phenylalanine 508 (refs 3, 4). Although the Δ F508 CFTR protein is in principle a functional anion channel, the protein is unstable and rapidly degraded, leading to an almost complete loss of CFTR channel function^{1,5–10}. Although both control (hereafter referred to as wild type, WT) and Δ F508 CFTR exhibit almost identical folds, the folding of Δ F508 CFTR is kinetically impaired, resulting in an increased recruitment of different chaperones¹¹. CF is therefore also characterized as a protein misfolding disease. Up to 90% of Δ F508 CFTR protein is retained in the endoplasmic reticulum (ER) and subsequently targeted for proteolytic degradation by the ER-associated degradation pathway (ERAD)^{8,10,12}. However, Δ F508 CFTR function can be partly rescued by a shift to lower temperature (26–30 °C)⁹ or inhibition of histone deacetylases (HDACi)^{13,14}. It is therefore likely that post-translational processes, such as altered chaperone recruitment, are critical for manifestation of CF. Accordingly, models have been proposed in which differential protein interactions with Δ F508 CFTR contribute to loss of function, but are favourably altered by temperature shift or HDACi¹¹. Yet relatively few proteins have been identified that interact with and participate in CFTR processing, in particular in bronchial epithelial cells, and it is largely unknown which interactions lead to Δ F508 CFTR stabilization and partial restoration of channel activity observed upon shift to permissive temperature or HDACi.

Δ F508 CFTR mutation-specific interactome

To identify interactions that potentially drive the disease phenotype, we developed co-purifying protein identification technology (CoPIT), an immunoprecipitation (IP)-based proteomic-profiling approach of protein–protein interactions across different sample conditions. Using CoPIT, which increased CFTR yield by 30- to 100-fold, we first determined the changes that occur between the WT and Δ F508 CFTR interactome in isogenic HBE41o– (WT CFTR) and CFBE41o– (Δ F508 CFTR) bronchial epithelial cell lines derived from a patient with CF¹⁵ (Fig. 1a and Extended Data Fig. 1). Proteins mapping to 638 genes were classified as high-confidence interactors. Δ F508 CFTR (Supplementary Data 1) and WT CFTR (Supplementary Data 2) interactomes comprised 576 and 430 proteins, respectively, with an overlap of more than 85% (Fig. 1b, c). These 638 proteins form the core CFTR interactome, and represent direct as well as indirect CFTR interactors (Supplementary Tables 1–3). An additional 915 interactors with medium confidence scores and at least a ratio of 10:1 over background were further assembled into an extended interactome (Extended Data Fig. 2a).

Although the majority of proteins (368) in the core interactome interact with both Δ F508 and WT CFTR, 209 differ significantly in the relative amounts recovered. An additional 208 and 62 interactors were detected only in Δ F508 CFTR and WT CFTR CoPIT experiments, respectively, which might represent interactors specific to or at least very highly enriched for either Δ F508 or WT CFTR. Protein expression profiling showed that the vast majority of observed differences between the Δ F508 and WT CFTR interactome are not due to altered expression levels of these proteins in the two cell lines (Extended Data Fig. 2b). Thus, a Δ F508 CFTR deletion-specific interactome was identified, which is characterized mainly by gain of novel interaction partners (Supplementary Table 5). Alterations in protein networks revealed distinct differences in the biogenesis of WT and Δ F508 CFTR.

¹Department of Chemical Physiology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. ²Department of Cell Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA.

*These authors contributed equally to this work.

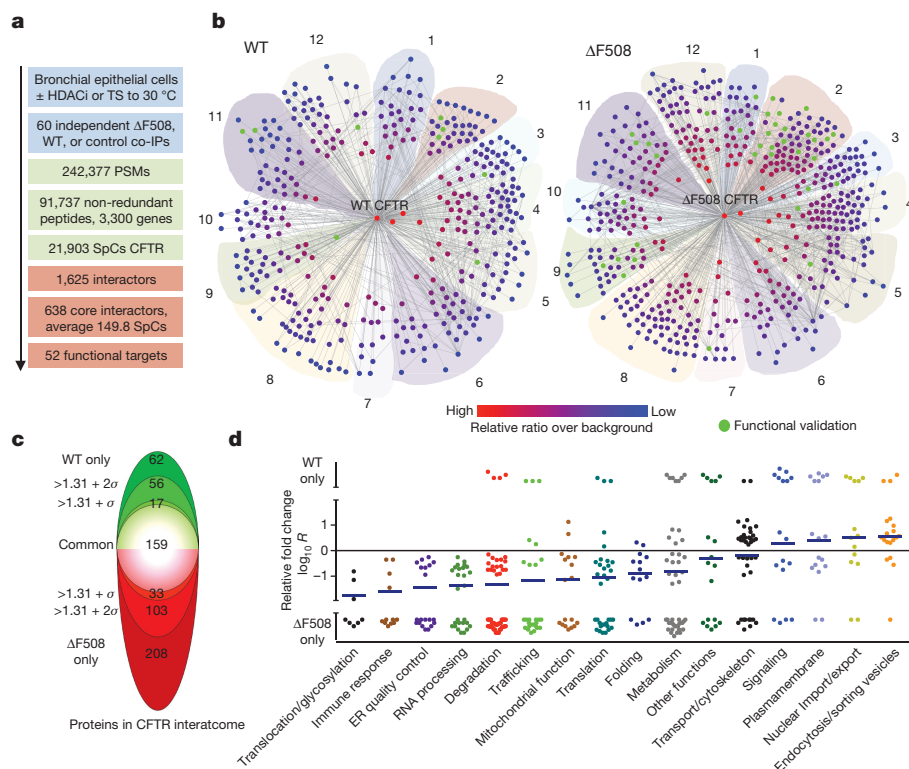


Figure 1 | WT and Δ F508 CFTR interactome in bronchial epithelial cells. **a**, Overview of workflow and results. **b**, Network representation of the WT and Δ F508 CFTR core interactome. Colour and distance to the centre (CFTR) reflect relative enrichment of individual interactors over background. Interactors targeted for functional rescue are in green (node labelling, Supplementary Figs 1 and 2). Proteins are grouped according to function: (1) protein folding, (2) protein degradation, ER quality control, (3) trafficking (4) protein transport, cytoskeleton, (5) endocytosis, plasma membrane micro-domain organization, (6) signalling, ion transport across membranes, (7) immune response, ROS signalling, (8) metabolism, lipid metabolism, mitochondrial function, (9) uncharacterized,

(10) DNA transcription, replication, repair, (11) RNA processing, nuclear import/export, (12) translation, post-translational modification, protein translocation. **c**, Venn diagram indicates the number of proteins significantly regulated between the WT (green) and Δ F508 CFTR (red) core interactome within different standard errors of measurement (σ) and those detected only in WT or Δ F508 CFTR-IPs. **d**, Plot depicts the top pathways affected by the Δ F508 mutation and individual regulation of identified CFTR interactors. Pathways are arranged in ascending order of the mean (blue horizontal line). Data are from independent biological replicates, Δ F508 CFTR $n = 8$, WT CFTR $n = 7$. TS, temperature shift; PSM, peptide spectrum match; SpC, spectral counts.

In particular, we observed enhanced recruitment of specific chaperones such as Hsp90 as well as enhanced protein degradation of Δ F508 CFTR mediated by a protein network, which increased vastly compared with the degradation and ER quality control network for WT CFTR and includes up to 25% of the Δ F508 CFTR specific interactions (Fig. 1d and Supplementary Table 6). While we recovered many of the proteins known to be involved in CFTR degradation, such as AMFR, STUB1 (CHIP) and VCP, we also identified several proteins that have been implicated previously in ERAD of other misfolded proteins but not of Δ F508 CFTR, including AUP1, SEL1L and FAF2 (ref. 16). Several of these novel interactions, such as with the lectin-binding protein LGALS3BP and the E3-ligase TRIM21, were confirmed by co-IP followed by western blot detection in bronchial epithelial cell lines and primary bronchial epithelial cells from patients with CF (Extended Data Fig. 2c–g). In addition, protein interactions implicated in translational control and messenger RNA (mRNA) decay, insertion of proteins into the ER (translocation), *N*-glycosylation, protein transport and trafficking, anchoring at the plasma membrane, as well as endocytic recycling were strongly altered, suggesting that the entire CFTR biogenesis is affected by deletion of F508. An example of such re-routing is the association of Δ F508 CFTR with the ER quality control component and sugar transferase UGGT, which re-glucosylates unfolded glycoproteins leading to eventual association with ERAD components, or the highly enhanced association of the co-chaperone PTPLAD1 with Δ F508 CFTR. Association of WT CFTR with components of Wnt and mTOR signalling pathways, and of Δ F508 CFTR with proteins involved in TGF- β and JAK/STAT signalling, suggests that cellular signalling is

also affected by the F508 deletion. Taken together, these data suggest that the loss of Δ F508 CFTR function emerges from novel associations with multiple alternative protein complexes and cellular pathways that route Δ F508 CFTR differently from WT CFTR.

Δ F508 CFTR interactome dynamics at 30 °C

Culture at 26–30 °C promotes formation of fully glycosylated Δ F508 CFTR (band C), incorporation into the plasma membrane and partial restoration of its channel activity⁹. To probe the temporal dynamics of interactions with Δ F508 CFTR and identify the molecular mechanisms that facilitate full glycosylation and lead to functional rescue of Δ F508 CFTR at lower temperature, we monitored changes of the Δ F508 CFTR interactome at different time points during temperature shift to 30 °C (Extended Data Fig. 3a). To this end, we first analysed the Δ F508 CFTR interactome by CoPIT after short (1 h, Supplementary Data 3), intermediate (6 h, Supplementary Data 4) and long (24 h, Supplementary Data 5) incubation at 30 °C, as well as upon reversal of the temperature shift (37 °C for 14 h after 24 h at 30 °C, Fig. 2a and Supplementary Data 6). Changes in the interactome were tightly coupled to the appearance of fully glycosylated Δ F508 CFTR (band C, Fig. 2b). Although few interactome changes were observed after 1 h at 30 °C, interactions with several proteins involved in ER quality control, such as AIMP1 and AUP1, and in lysosomal targeting (LAMP1) were reduced, and a few new interactions were gained (Supplementary Table 7). Long-term incubation at 30 °C abolished 186 (89%) of the 208 unique interactions (Fig. 2c) and the interactome was extensively remodelled with more than 65% of all interactions altered. The increased presence of

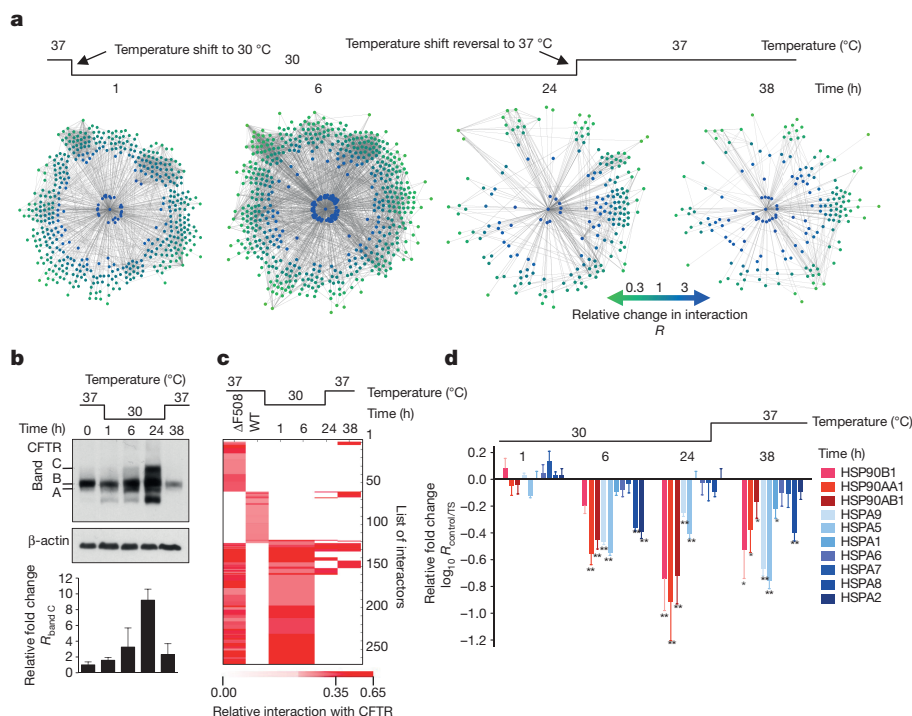


Figure 2 | Dynamic changes of the $\Delta F508$ CFTR interactome during temperature shift to 30 °C. **a**, Network representation of $\Delta F508$ CFTR interactome changes occurring at different time points during temperature shift. Colour and distance to CFTR (centre node) indicate fold change of individual interactors (green, reduced association; blue, enhanced association). The innermost 'circle' contains interactors gained during temperature shift (node labelling, Supplementary Figs 3–6). Proteins are grouped according to function as in Fig. 1b. **b**, Western blot showing the effect of the temperature shift on non-glycosylated (band A),

band C was reflected in the interactome as follows: first, by reduced association of $\Delta F508$ CFTR with degradation promoting proteins of ubiquitin-mediated pathways and ERAD, as well as of those involved in endocytic removal of plasma membrane proteins; second, by a more favourable folding environment marked by decreased recruitment of heat-shock proteins (Fig. 2d); and third, by a marked downregulation of RNA processing (including mRNA decay) proteins such as PABPC1 (33-fold) (Supplementary Tables 7–10).

Reversal of the temperature shift led to loss of fully glycosylated $\Delta F508$ CFTR. However, with only 20 $\Delta F508$ CFTR-specific interactions re-established, the interaction profile still clustered with that of WT CFTR. Interactions that mediated CFTR degradation from either the cell surface or ER, such as E3-ubiquitin protein ligases AMFR (gp78) and STUB1 (refs 17–19), were re-gained first. Association of $\Delta F508$ CFTR with RAB5B and RAB5A, which are involved in apical endocytosis and recycling, as well as with Erlin1 and Erlin2, which have been implicated in ERAD of IP3 receptors²⁰, was also restored. The experiment thus indicated that removal from the plasma membrane and subsequent degradation as well as degradation of newly synthesized $\Delta F508$ CFTR in the ER is responsible for the rapid loss of fully glycosylated $\Delta F508$ CFTR. Taken together, the temperature shift experiment revealed that the association of $\Delta F508$ CFTR with the mutation-specific interactome and consequent alteration of CFTR biogenesis can be suppressed by temperature shift and thus may be responsible for the functional rescue.

Interactome remodelling upon HDACi

Recently, it was reported that inhibition of HDAC activity leads to increased presence of fully glycosylated $\Delta F508$ CFTR and partial functional rescue¹³. Monitoring the interactome upon short interfering RNA (siRNA)-mediated knockdown of HDAC7 (Supplementary Data 7),

core-glycosylated (band B) and fully glycosylated $\Delta F508$ CFTR (band C). Bar graph displays induction of band C during temperature shift relative to control (0 h). Error bars, mean \pm s.e.m. **c**, Heat map of temperature-sensitive $\Delta F508$ CFTR and WT CFTR-specific interactions. Colour represents protein abundance relative to CFTR. **d**, Differential interactions of heat-shock proteins (HSPs) with $\Delta F508$ CFTR over time during temperature rescue and reversal. Data represent independent biological replicates, WT ($n = 7$), $\Delta F508$ ($n = 8$), 1 h ($n = 4$), 6 h ($n = 4$), 24 h ($n = 2$), 24 h reversed ($n = 2$). See Methods for statistical analysis.

or treatment with 100 nM trichostatin A (TSA) (Supplementary Data 8) or 5 μ M suberoylanilide hydroxamic acid (SAHA) (Supplementary Data 9) for 24 h, revealed that HDACi induced similar large-scale changes to the $\Delta F508$ CFTR interactome as the temperature shift (Fig. 3a, Supplementary Tables 11–13, Supplementary Results and Discussion). More than 75% and almost 90% of interactions affected by TSA or HDAC7 knockdown were also altered by SAHA treatment (Fig. 3b). In particular, HDACi abolished interactions that were either specific or preferential for $\Delta F508$ CFTR and restored a few WT CFTR-specific interactions (Fig. 3c), such as with the proteins NHERF1 and NHERF2, which can act as apical plasma membrane adapters for WT CFTR, and thus probably reflect enhanced $\Delta F508$ CFTR stability at the plasma membrane.

Comparison of the interactions that were affected by temperature shift and HDACi identified trafficking, degradation and mRNA decay pathways required for $\Delta F508$ CFTR rescue and pinpointed distinct differences in the mechanisms by which $\Delta F508$ CFTR rescue is achieved. In contrast to temperature shift, TSA failed to reduce association with several protein disulfide isomerases that are involved in ER quality control. SAHA treatment even enhanced association with ERAD component SEL1L, with E3-ligase SUGT1 and E3C ligase (UBE3C), which enhances proteasome processivity²¹. We also identified additional lysosomal degradation proteins such as cathepsin B and TPP1 in the SAHA interactome, probably reflecting failure of SAHA to fully prevent retrotranslocation and degradation of $\Delta F508$ CFTR. Additionally, HDACi induced extensive changes to the $\Delta F508$ CFTR-associated cytoskeleton, which appear to have wide-ranging influence on anterograde and retrograde transport. Despite these changes to the interactome, the interaction profiles of $\Delta F508$ CFTR upon treatment with HDACi or Cmpd 4a clustered with the interaction profile of control $\Delta F508$ CFTR rather than with WT CFTR (Extended Data Fig. 3b). Further differences

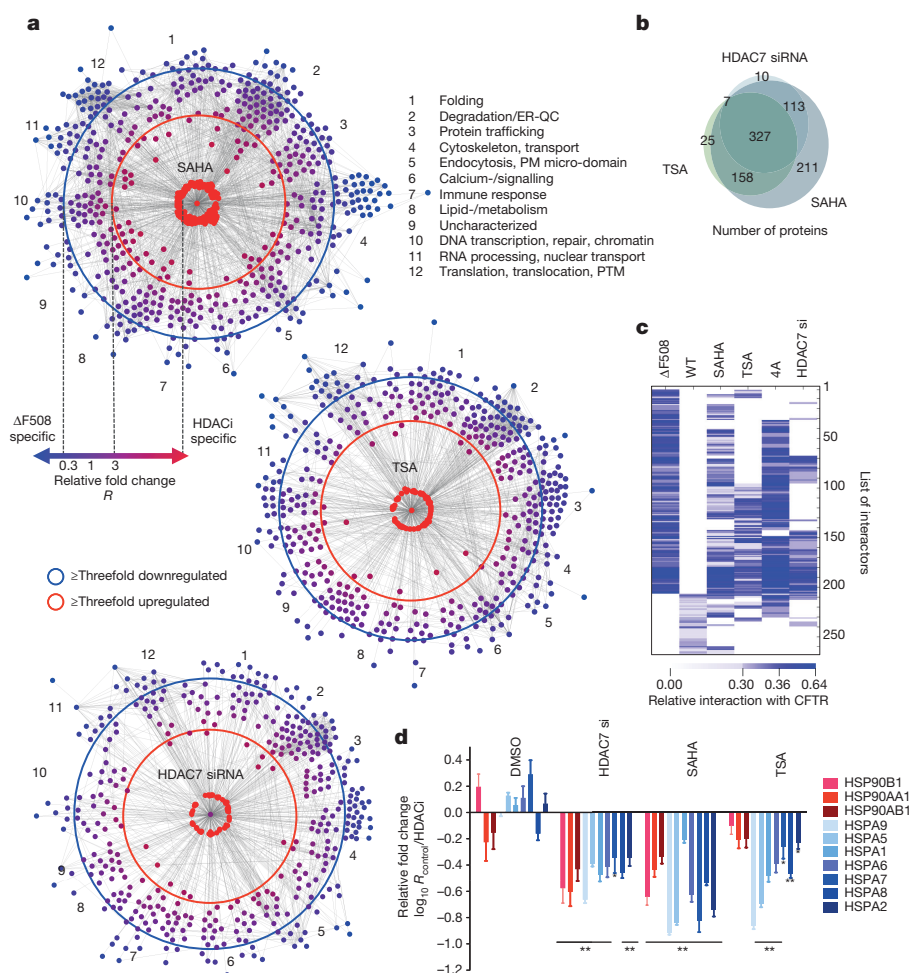


Figure 3 | HDACi sensitive changes of the $\Delta F508$ CFTR interactome.

a, Network representation of dynamic changes in the $\Delta F508$ CFTR interactome upon HDAC7 knockdown or treatment with SAHA or TSA. Distance to $\Delta F508$ CFTR and colour represent fold change of individual interactors (blue, reduced; red, enhanced, node labelling; Supplementary Figs 7–9). Proteins are grouped according to function. **b**, Heat map of HDACi sensitive $\Delta F508$ and WT CFTR-specific interactions. $\Delta F508$ CFTR co-IP results from CFBE41o⁻ cells treated with 10 μM of

Cmpd 4a (CFF) are included for comparison. **c**, Proportional Venn diagram depicting the overlap of $\Delta F508$ CFTR interactions affected by SAHA, TSA or HDAC7 siRNA. **d**, Differential interactions of heat-shock proteins (HSPs) with $\Delta F508$ CFTR upon treatment with SAHA, TSA or dimethylsulfoxide (DMSO; control) or HDAC7 knockdown. All data represent independent biological replicates, WT ($n = 7$), $\Delta F508$ ($n = 8$), SAHA ($n = 4$), TSA ($n = 4$), HDAC7 si ($n = 3$), Cmpd 4a ($n = 3$). See Methods for statistical analysis.

between temperature-shift- and HDACi-mediated rescue included an inversely altered association of chaperone HSP70 and HSP90 family members with $\Delta F508$ CFTR (Fig. 3d). While temperature shift only slightly affected association of $\Delta F508$ CFTR with the HSP70 and Hsc70 chaperone machinery (1.35-fold less), it strongly reduced the association of $\Delta F508$ CFTR with Hsp90 proteins (6.2-fold less). Conversely, HDACi strongly reduced the association of $\Delta F508$ CFTR with detected Hsp70 family members (3.4-fold less) and affected association with Hsp90 proteins to a lesser degree (2.5-fold less). Reduced binding of chaperones to $\Delta F508$ CFTR was independent of chaperone expression levels, which were either not influenced or upregulated by temperature shift or HDACi¹³ (Extended Data Fig. 3c). However, enhanced acetylation of heat-shock proteins may lead to remodelling of the chaperone environment and may disrupt the heat-shock–ubiquitin–proteasome pathway, which controls mRNA decay²². $\Delta F508$ CFTR mRNA decay is possibly the pacemaker for the CF phenotype, as all treatments that induced $\Delta F508$ CFTR rescue downregulate the association of a distinct set of more than 30 proteins that affect mRNA stabilization and decay, including PABPC1, YBX1 and UPF1.

Interestingly, a subset of seven $\Delta F508$ CFTR-specific interactions was not corrected either by temperature shift or by SAHA. This subset included members of the 26S proteasome (PSMC1, PSMD11), which induce protein aggregation and neuro-degeneration if inhibited

in their function^{23,24}, and PSMB8, a stress-inducible subunit of the 20S core proteasome²⁵, as well as the two co-chaperones BAG3 and DNAJB2. DNAJB2 inhibition leads to partial $\Delta F508$ CFTR rescue¹⁹. BAG3, whose binding to $\Delta F508$ CFTR was significantly upregulated immediately after temperature shift, mediates aggresome formation and selectively induces autophagy of misfolded proteins²⁶. Persistence of these interactions suggests that these proteins detect $\Delta F508$ CFTR and channel it to autophagy and proteasomal degradation, even under rescuing conditions. SURF4 has been implicated in vesicular trafficking^{27,28} and store-operated Ca^{2+} entry²⁹, whereas the molecular function of the last member of this subset, ERH, has remained enigmatic, but may be associated with RNA splicing³⁰.

Interactor RNA interference restores $\Delta F508$ function

To assess the potential of rescuing the $\Delta F508$ CFTR phenotype by blocking novel protein–protein interactions identified in this study, we performed an RNA interference (RNAi) screen with validated short hairpin RNAs (shRNAs) and monitored $\Delta F508$ CFTR maturation and its glycosylation pattern by electrophoresis as a measure of rescue. A total of 52 proteins were selected (Extended Data Fig. 4) and tested including HDAC2 as positive and CSNK2A as negative controls. Knockdown of 6 proteins had minor to no effect, knockdown of 17 proteins led to reduced $\Delta F508$ CFTR stability and yield,

and knockdown of 31 interactors promoted Δ F508 CFTR maturation (Fig. 4a and Extended Data Fig. 5). Many of the 31 novel interactors might sequentially control Δ F508 CFTR protein production and turn over as they belong to (1) a network associated with mRNA decay and co-translational control, (2) complexes affecting Δ F508 CFTR trafficking and endocytic recycling, (3) ER quality control and folding or (4) the protein degradation network (Fig. 4b; see also Supplementary Results and Discussion). The subcellular interaction of Δ F508 CFTR with the top sub-networks or complexes was spatially resolved by co-immunostainings of nine binding partners that represent different cellular compartments according to Gene Ontology (Fig. 4c and Extended Data Fig. 6a–c). Prolyl-4-hydroxylase (P4HB), an ER and plasma membrane marker, PDIA4, which recognizes unfolded protein regions³¹, and PTPLAD1, which exhibits Hsp90 co-chaperone activity³², co-localized with Δ F508 CFTR in the ER. Co-staining was also observed with SURF4, which is found in the early secretory pathway, ERGIC and Golgi²⁸, as well as with the GTPase RASEF, which is potentially involved in membrane trafficking. Co-staining of Δ F508 CFTR with KLHDC10 and TRIM21, which are involved in degradation^{33,34}, and with PABPC1, which is involved in RNA processing^{35–37}, was observed in the nuclear periphery. LGALS3BP,

which is part of the KLHDC10–FAF2 degradation complex³⁸ and which negatively influenced Δ F508 CFTR stability, only partly co-localized with Δ F508 CFTR in vesicular structures.

To evaluate further the therapeutic potential of interactors that influenced Δ F508 CFTR maturation in CFBE41o– cells in the RNAi screen, we assessed rescue of Δ F508 CFTR channel function for eight interactors that bind preferentially to Δ F508 CFTR and/or were dynamically regulated by temperature shift and HDACi. Interactors represent either the RNA decay and co-translational control network (PABPC1, PTBP1, YBX1), or the degradation network (LGALS3BP, TRIM21) or are potential novel components of ER quality control (PDIA4, SURF4, PTPLAD1). Primary human bronchial epithelial cells from healthy donors or patients with CF, and CFBE41o– cells, were differentiated into epithelial cultures at an air–liquid interface (ALI) and Δ F508 CFTR channel function was determined by electrophysiology in an Ussing chamber (Fig. 5a).

Knockdown of seven interactors enhanced forskolin/genistein-stimulated Δ F508 CFTR channel activity at the apical plasma membrane up to 12-fold over controls in primary CF epithelia and by about 4.5- to 7-fold in CFBE41o– epithelia, which is comparable to rescue by temperature shift (Fig. 5b, c). As determined by western

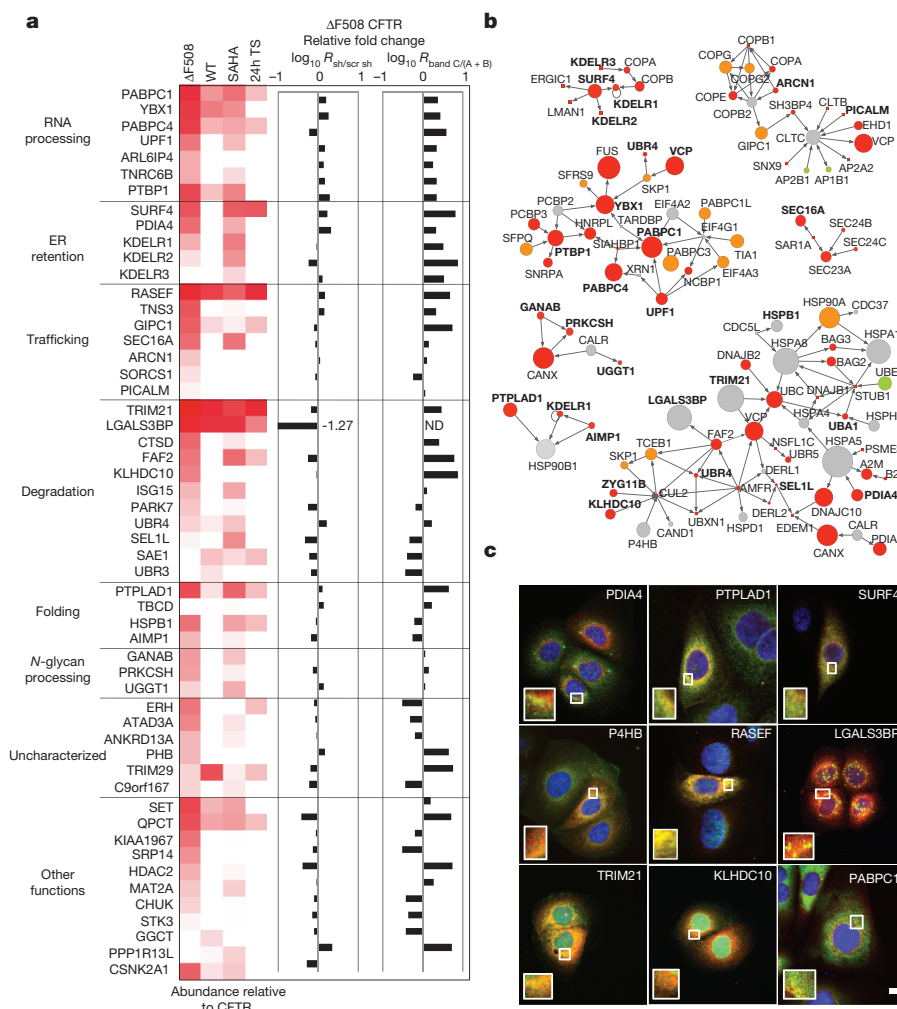


Figure 4 | RNAi and subnetworks of novel key interactors. **a**, Heat map indicating relative abundance of interactors selected for RNAi (left; white, absent; red, highly abundant). Bar graph shows the relative change of total Δ F508 CFTR protein (central) and of the ratio of Δ F508 CFTR band C to (band A + band B) as determined by western blot (right) upon RNAi. ND, value not determined. Data are representative of at least two independent knockdown experiments per target protein. **b**, RNAi candidate sub-networks in the CFTR interactome (bold). Colouring indicates relative

fold change and significance (Δ F508/WT CFTR): red, enhanced ($\geq 2\sigma$); orange, enhanced ($\geq 1\sigma$); green, decreased ($\geq 2\sigma$); light green, decreased ($\geq 1\sigma$); grey, non-significant. Node size reflects \log_{10} (interactor abundance) in Δ F508 CFTR IPs. **c**, Co-localization of Δ F508 CFTR (red) with select interactors (green) representing different sub-networks and complexes. Nuclei were counterstained with 4',6-diamidino-2-phenylindole (DAPI) (blue). Boxed areas are magnified in insets (DAPI stain excluded). Scale bar, 10 μ m.

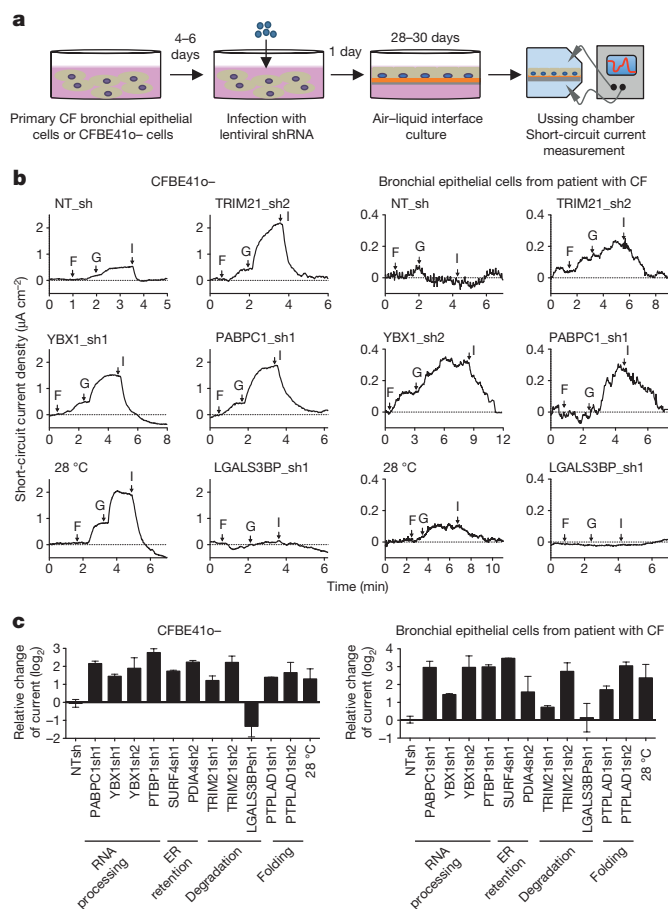


Figure 5 | Rescue of Δ F508 CFTR channel function defect by knockdown of Δ F508 CFTR interactors in human primary CF bronchial epithelial cells and CFBE410 $^{-}$ cells. a, Experiment setup. Primary bronchial epithelial cells or CFBE410 $^{-}$ cells were infected with shRNA lentivirus, seeded onto Snapwell culture inserts and cultured at air-liquid interface for 28–30 days, before measuring short-circuit currents in an Ussing chamber. **b**, Representative traces of forskolin (10 μ M) and genistein (50 μ M) activated Δ F508 CFTR short-circuit current (I_{sc}). **c**, Quantification of the peak CFTR Inhibitor 172 (Inh 172)-sensitive I_{sc} (ΔI_{sc}) in CFBE410 $^{-}$ cells ($n = 3-5$) and in human primary CF bronchial epithelial cells (DHBE, $n = 2-5$) as fold change relative to non-target shRNA (NT sh) following knockdown of indicated interactors. Error bars, mean \pm s.e.m.

blot, knockdown of seven of the eight interactors also led to a clearly visible Δ F508 CFTR signal in the primary ALI cultures after differentiation for 28 d and induced band C formation similar to temperature shift, which correlates well with the increase in Δ F508 CFTR channel activity observed in the Ussing chamber measurements (Extended Data Fig. 7). In the case of LGALS3BP knockdown, no CFTR signal was detected in primary CF bronchial epithelial cells by western blot and we failed to detect Δ F508 CFTR-specific chloride current in CFBE410 $^{-}$ epithelia or primary CF bronchial epithelia. Loss of Δ F508 CFTR in CFBE410 $^{-}$ cells that constitutively express an LGALS3BP shRNA (clone 13) showed that LGALS3BP is critical for Δ F508 CFTR stability. Furthermore, no CFTR chloride channel activity was measured upon LGALS3BP knockdown in a halide-sensitive yellow fluorescent protein (YFP) assay, whereas upon stable knockdown of PTPLAD1 (clone 24), CFTR chloride channel function was greater than in parental CFBE410 $^{-}$ cells (Extended Data Fig. 8). Our results show that reduction of protein levels of seven interactors rescues channel function of Δ F508 CFTR and thus we conclude that modulation of interactors might be a promising route to correction of the Δ F508 CFTR defect.

Closing remarks and outlook

The CoPIT results established a comprehensive interactome for WT as well as Δ F508 CFTR in epithelial airway cells, defined disease-specific alterations and revealed interactome dynamics upon temperature shift and intervention by HDACi. The number of proteins obtained for the CFTR core interactome with CoPIT (638) can be rationalized by the identification of direct and indirect interactors of CFTR (second- and third-degree interactions) and reflects the complicated multi-step biogenesis of membrane proteins in mammalian cells as well as the number of different possibilities of a cell to cope with misfolded CFTR protein. Δ F508 alters CFTR translation, folding, insertion into the ER and trafficking, and enhances its degradation, overall contributing to an increased number of direct and indirect interactors compared with WT CFTR. Thus, CoPIT analysis of the CFTR interactome shows that the disease phenotype CF is a direct consequence of the derailment of a whole network of protein interactions in the presence of the Δ F508 mutation.

Intriguingly, many of the proteins that bind differentially to WT and Δ F508 CFTR have been implicated in other misfolding or protein aggregation diseases as well, as revealed by querying the Online Mendelian Inheritance in Man (OMIM) database³⁹ and UniprotKB⁴⁰. In particular, we noticed differential binding of proteins to CFTR that are implicated in neurodegenerative diseases (Extended Data Fig. 9), suggesting similar disease mechanisms. Although we can only speculate, the mechanisms that lead to Δ F508 CFTR destabilization and clearance could be tentatively harvested to achieve clearance of toxic protein aggregates or to stabilize other misfolded proteins that display a loss of function phenotype such as Δ F508 CFTR.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 September 2013; accepted 14 September 2015.

Published online 30 November 2015.

1. Dalemans, W. *et al.* Altered chloride ion channel kinetics associated with the Δ F508 cystic fibrosis mutation. *Nature* **354**, 526–528 (1991).
2. Collins, F. S. Cystic fibrosis: molecular biology and therapeutic implications. *Science* **256**, 774–779 (1992).
3. Watson, M. S. *et al.* Cystic fibrosis population carrier screening: 2004 revision of American College of Medical Genetics mutation panel. *Genet. Med.* **6**, 387–391 (2004).
4. World Health Organization. *The Molecular Genetic Epidemiology of Cystic Fibrosis* http://www.who.int/genomics/publications/en/HGN_WB_04.02_report.pdf (World Health Organization, 2004).
5. Lukacs, G. L. *et al.* The Δ F508 mutation decreases the stability of cystic fibrosis transmembrane conductance regulator in the plasma membrane. Determination of functional half-lives on transfected cells. *J. Biol. Chem.* **268**, 21592–21598 (1993).
6. Drumm, M. L. *et al.* Chloride conductance expressed by Δ F508 and other mutant CFTRs in *Xenopus* oocytes. *Science* **254**, 1797–1799 (1991).
7. Li, C. *et al.* The cystic fibrosis mutation (Δ F508) does not influence the chloride channel activity of CFTR. *Nature Genet.* **3**, 311–316 (1993).
8. Lukacs, G. L. *et al.* Conformational maturation of CFTR but not its mutant counterpart (Δ F508) occurs in the endoplasmic reticulum and requires ATP. *EMBO J.* **13**, 6076–6086 (1994).
9. Denning, G. M. *et al.* Processing of mutant cystic fibrosis transmembrane conductance regulator is temperature-sensitive. *Nature* **358**, 761–764 (1992).
10. Jensen, T. J. *et al.* Multiple proteolytic systems, including the proteasome, contribute to CFTR processing. *Cell* **83**, 129–135 (1995).
11. Wang, X. *et al.* Hsp90 cochaperone Aha1 downregulation rescues misfolding of CFTR in cystic fibrosis. *Cell* **127**, 803–815 (2006).
12. Ward, C. L., Omura, S. & Kopito, R. R. Degradation of CFTR by the ubiquitin-proteasome pathway. *Cell* **83**, 121–127 (1995).
13. Hutt, D. M. *et al.* Reduced histone deacetylase 7 activity restores function to misfolded CFTR in cystic fibrosis. *Nature Chem. Biol.* **6**, 25–33 (2010).
14. Boyault, C. *et al.* HDAC6-p97/VCP controlled polyubiquitin chain turnover. *EMBO J.* **25**, 3357–3366 (2006).
15. Bruscia, E. *et al.* Isolation of CF cell lines corrected at Δ F508-CFTR locus by SFHR-mediated targeting. *Gene Ther.* **9**, 683–685 (2002).
16. Mueller, B., Klemm, E. J., Spooner, E., Claessen, J. H. & Ploegh, H. L. SEL1L nucleates a protein complex required for dislocation of misfolded glycoproteins. *Proc. Natl Acad. Sci. USA* **105**, 12325–12330 (2008).

17. Vij, N., Fang, S. & Zeitlin, P. L. Selective inhibition of endoplasmic reticulum-associated degradation rescues Δ F508-cystic fibrosis transmembrane regulator and suppresses interleukin-8 levels: therapeutic implications. *J. Biol. Chem.* **281**, 17369–17378 (2006).
18. Meacham, G. C., Patterson, C., Zhang, W., Younger, J. M. & Cyr, D. M. The Hsc70 co-chaperone CHIP targets immature CFTR for proteasomal degradation. *Nature Cell Biol.* **3**, 100–105 (2001).
19. Okiyonedo, T. *et al.* Peripheral protein quality control removes unfolded CFTR from the plasma membrane. *Science* **329**, 805–810 (2010).
20. Pearce, M. M., Wormer, D. B., Wilkens, S. & Wojcikiewicz, R. J. An endoplasmic reticulum (ER) membrane complex composed of SPFH1 and SPFH2 mediates the ER-associated degradation of inositol 1,4,5-trisphosphate receptors. *J. Biol. Chem.* **284**, 10433–10445 (2009).
21. Chu, B. W. *et al.* The E3 ubiquitin ligase UBE3C enhances proteasome processivity by ubiquitinating partially proteolyzed substrates. *J. Biol. Chem.* **288**, 34575–34587 (2013).
22. Laroia, G., Cuesta, R., Brewer, G. & Schneider, R. J. Control of mRNA decay by heat shock-ubiquitin-proteasome pathway. *Science* **284**, 499–502 (1999).
23. Bedford, L. *et al.* Depletion of 26S proteasomes in mouse brain neurons causes neurodegeneration and Lewy-like inclusions resembling human pale bodies. *J. Neurosci.* **28**, 8189–8198 (2008).
24. Santamaria, P. G., Finley, D., Ballesta, J. P. & Remacha, M. Rpn6p, a proteasome subunit from *Saccharomyces cerevisiae*, is essential for the assembly and activity of the 26 S proteasome. *J. Biol. Chem.* **278**, 6687–6695 (2003).
25. Callahan, M. K., Wohlfert, E. A., Menoret, A. & Srivastava, P. K. Heat-shock up-regulates Imp2 and Imp7 and enhances presentation of immunoproteasome-dependent epitopes. *J. Immunol.* **177**, 8393–8399 (2006).
26. Gamerdinger, M., Kaya, A. M., Wolfrum, U., Clement, A. M. & Behl, C. BAG3 mediates chaperone-based aggresome-targeting and selective autophagy of misfolded proteins. *EMBO Rep.* **12**, 149–156 (2011).
27. Matsuzaki, F., Shirane, M., Matsumoto, M. & Nakayama, K. I. Protrudin serves as an adaptor molecule that connects KIF5 and its cargoes in vesicular transport during process formation. *Mol. Biol. Cell* **22**, 4602–4620 (2011).
28. Mitrovic, S., Ben-Tekaya, H., Koegler, E., Gruenberg, J. & Hauri, H. P. The cargo receptors Surf4, endoplasmic reticulum-Golgi intermediate compartment (ERGIC)-53, and p25 are required to maintain the architecture of ERGIC and Golgi. *Mol. Biol. Cell* **19**, 1976–1990 (2008).
29. Fujii, Y. *et al.* Surf4 modulates STIM1-dependent calcium entry. *Biochem. Biophys. Res. Commun.* **422**, 615–620 (2012).
30. Weng, M. T. & Luo, J. The enigmatic ERH protein: its role in cell cycle, RNA splicing and cancer. *Protein Cell* **4**, 807–812 (2013).
31. Forster, M. L. *et al.* Protein disulfide isomerase-like proteins play opposing roles during retrotranslocation. *J. Cell Biol.* **173**, 853–859 (2006).
32. Taguwa, S. *et al.* Cochaperone activity of human butyrate-induced transcript 1 facilitates hepatitis C virus replication through an Hsp90-dependent pathway. *J. Virol.* **83**, 10427–10436 (2009).
33. Mallery, D. L. *et al.* Antibodies mediate intracellular immunity through tripartite motif-containing 21 (TRIM21). *Proc. Natl Acad. Sci. USA* **107**, 19985–19990 (2010).
34. Behrends, C., Sowa, M. E., Gygi, S. P. & Harper, J. W. Network organization of the human autophagy system. *Nature* **466**, 68–76 (2010).
35. Uchida, N., Hoshino, S. & Katada, T. Identification of a human cytoplasmic poly(A) nuclease complex stimulated by poly(A)-binding protein. *J. Biol. Chem.* **279**, 1383–1391 (2004).
36. Fabian, M. R. *et al.* Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation. *Mol. Cell* **35**, 868–880 (2009).
37. LaCava, J. *et al.* RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **121**, 713–724 (2005).
38. Alexandru, G. *et al.* UBXD7 binds multiple ubiquitin ligases and implicates p97 in HIF1 α turnover. *Cell* **134**, 804–816 (2008).
39. McKusick, V. A. Mendelian Inheritance in Man and its online version, OMIM. *Am. J. Hum. Genet.* **80**, 588–604 (2007).
40. Magrane, M. & UniProt Consortium. UniProt Knowledgebase: a hub of integrated protein data. *Database* **2011**, bar009 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Bamberger for many discussions on statistical analysis, and D. Cociorva and T. Xu for making DTASelect 2.1 available and for suggestions and comments on data analysis strategies. We further thank D. Hutt and D. Roth for discussion. This work was supported by National Institutes of Health grants 5R01HL079442-08 (to J.R.Y. and W.E.B.), P01AG031097 (to J.R.Y.), P41 GM103533 (to J.R.Y.), HHSN268201000035C (to J.R.Y.), and a Cystic Fibrosis Foundation mass spectrometry fellowship BALCH050X6 (to S.P. and J.R.Y.). M.L.-A. holds a postdoctoral fellowship from Fonds de recherche du Québec - Nature et technologies. W.E.B. is supported by National Institutes of Health grants R01HL095524 and R01DK051870 and Tobacco-Related Disease Research Program grant 23TRU0012.

Author Contributions Analysed data are available at <http://sealion.scripps.edu/pint?project=CFTR> ('CFTR' dataset). Raw files have been deposited in ProteomeXchange under accession number PXD002722. S.P. and C.B. developed experimental methods and performed all experiments, data analysis and mass spectrometric (MS) measurements. C.B. and S.P. developed and performed statistical analysis. D.C. and C.B. developed Radial Topology Viewer. M.L.-A. performed and wrote the methods detailing the hierarchical clustering. S.M.B. developed and maintains Proteomics INTEgrator (PINT). J.R.Y., W.E.B., S.P. and C.B. designed the study. S.P., C.B. and J.R.Y. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.R.Y. (jyates@scripps.edu) or S.P. (pankows@scripps.edu).

METHODS

Cell lines and cell culture. Human bronchial epithelial cells (CFBE41o–) carrying the Δ F508 CFTR mutation, or HBE41o– cells harbouring a WT CFTR allele, and isogenic CFTR null cells (CFBE41o–, null) were provided by J. Clancy. Cells were cultured at 37°C, 5% CO₂ in Advanced MEM (Gibco) supplemented with 1% penicillin/streptomycin (Gibco), 10% fetal bovine serum (Gibco) and 2 mM L-glutamine (Gibco) and appropriate selective antibiotics. Twenty hours before IP, cells were treated with DMSO (vehicle) or DMSO and 100 nM TSA (Sigma-Aldrich), 5 μ M SAHA (Cayman Chemicals), 15 μ M N-[2-(5-chloro-2-methoxyphenylamino)-4'-methyl-[4,5']bithiazolyl-2'-yl]-benzamide (Cmpd 4a, C4, Cystic Fibrosis Foundation, <http://www.cftrfolding.org/CFTRReagents.htm>). For siRNA-mediated knock down of HDAC7, CFBE41o– cells were transfected with Lipofectamine RNAiMAX (Invitrogen) and 50 nM of validated HDAC7-specific siRNA (Ambion) or scrambled control siRNA (Ambion) according to the manufacturers' protocol. The cell culture medium was changed the next day and cells harvested 72 h after transfection. Primary bronchial epithelial cells were obtained from the Cystic Fibrosis Center at the University of Alabama according to institutional review board regulations or from Lonza, and were cultured in complete BEGM medium (Lonza) at 37°C, 5% CO₂ for up to three passages, starting with passage 0. Cell lines were tested for mycoplasma contamination with DAPI staining and their identity validated based on presence of the correct CFTR alleles as determined with polymerase chain reaction (PCR) and mass spectrometry. Cell culture experiments as well as subsequent sample processing and mass spectrometric data-taking were randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Lentiviral-mediated knockdown of target proteins. Lentiviral particles containing shRNA sequences specific for the target proteins were generated in HEK293T cells using the Mission shRNA system with validated shRNA sequences (Sigma-Aldrich) following standard protocols⁴¹. CFBE41o– cells were infected with lentiviral particles for 16 h and cultivated for additional 48 h before harvest. Lentivirus production and infection is covered under approval 01-13-10-07 from The Scripps Research Institute and all steps were performed in a biosafety level 2/3-certified laboratory. Rescue of Δ F508 CFTR was monitored by western blotting followed by immunodetection of CFTR using rat monoclonal 3G11 antibody. The RNAi Consortium identification numbers for the shRNAs used are given in Supplementary Table 15.

Western blotting and immunofluorescence. Protein lysates were prepared as described above, denatured in SDS sample buffer⁴² either for 15 min at 37°C to detect CFTR or for 5 min at 95°C, separated by SDS–polyacrylamide gel electrophoresis and transferred onto nitrocellulose (Protran; Schleicher & Schuell). The following primary antibodies were used: rat monoclonal antibody against CFTR (3G11), mouse monoclonal antibodies against CFTR (24.1, ATCC; M3A7, Chemicon) and β -actin (AC-15, Sigma), rabbit polyclonal antibodies against HDAC2 (9928S, Cell Signaling), PABPC1 (ab21060, Abcam), anti-galectin-3BP (AF2226, R&D Systems), anti-PTPLAD1 (WH0051495M1, Sigma), anti-52 kDa Ro/SSA (sc-25351, Santa Cruz) and anti-Na⁺/K⁺ATPase α Antibody (H-300, sc28800, Santa Cruz). Horseradish-peroxidase-conjugated secondary antibodies (Jackson ImmunoResearch) were detected with enhanced chemiluminescence reagent (ECL, Pierce). For immunofluorescence, CFBE41o– cells fixed with 4% paraformaldehyde were permeabilized with 0.1% Triton X100, blocked in 10% FBS in 1 \times PBS for 1 h at room temperature (21°C) and incubated with the following antibodies for 4 h at room temperature (21°C): anti-CFTR (3G11), anti-galectin-3BP (R&D Systems, AF2226), anti-PTPLAD1 (Sigma, WH0051495M1), anti-KLHDC10 (Sigma, HPA020119), anti-52 kDa Ro/SSA (Santa Cruz, sc-25351), anti-Rab45 (Santa Cruz, sc-81925), anti-Surfeit4 (Santa Cruz, sc-107304), anti-Erp72 (Abcam, ab82587, Enzo ADI-SPS-720), anti-PABPC1 (Abcam, ab21060) and anti-P4HB (3501S, Cell Signaling). AlexaFluor 488-, DyLight 488- or DyLight 549-conjugated secondary antibodies (Jackson ImmunoResearch) were used for detection of the primary antibodies. Nuclei were counterstained with DAPI (Molecular Probes, Invitrogen). Photographs of cells mounted in ProLong Gold antifade reagent (Molecular Probes, Invitrogen) were taken with a laser scanning confocal microscope LSM 710 (Zeiss) or Radiance 2100 Rainbow (Zeiss).

Premo Halide Sensor assay. Chloride channel activity of CFTR was determined with a Premo Halide Sensor assay (Invitrogen) measuring quenching of a halide-sensitive YFP variant (Venus YFP). To this end, HBE41o–, CFBE41o– and CFBE41o– cell lines stably transduced with shRNA lentivirus were infected with the Bacman gene delivery system to introduce the halide-sensitive Venus YFP. Subsequently, cells were seeded in glass bottom 96- or 24-well plates and cultivated overnight. Quenching of YFP fluorescence by iodide influx was measured at single-cell level with a Radiance 2100 Rainbow laser scanning confocal microscope (Zeiss) according to the protocol initially established in ref. 43. Briefly, before time-lapse recording, cells were pre-incubated with 50 μ M genistein. Cells with sufficient

YFP fluorescence were then selected and data acquisition was started with a frame speed of 0.5–1.0 s. After 5 s, sodium iodide was added to a final concentration of 0.1 M and chloride channel activity was further stimulated by addition of forskolin (20 μ M). Acquired data were analysed with Matlab (<http://www.mathworks.com>) and Prism (GraphPad Software), and decay curves were fitted over the time course. At least ten individual cells for each cell line were recorded per experiment.

Ussing chamber measurements. Primary human CF and control (WT) bronchial epithelial cells infected with Mission shRNA lentiviral particles with a multiplicity of infection between 3 and 5 were plated on 12 mm Snapwell membranes (Corning) coated with rat tail collagen I (BD Biosciences) at a density of 1×10^5 cells per square centimetre and cultured in BEGM. Upon confluence, cells were maintained in B-ALI differentiation medium (Lonza) under ALI conditions for at least 21 d. Transepithelial resistance (R_T) of the ALI cultures was measured with a Millicell ERS2 Volt ohmmeter (Millipore) and was between 200 and 2,700 Ω cm^{–2}. Polarized cultures were mounted in EasyMount Ussing chambers (Physiological Instruments), bathed bilaterally with Krebs–Ringer bicarbonate solution (140 mM Na⁺, 119.8 mM Cl[–], 25 mM HCO₃[–], 2.8 mM K⁺, 2.4 mM HPO₄^{2–}, 0.4 mM PO₄^{3–}, 1.2 mM Mg²⁺, 1.2 mM Ca²⁺, 5 mM glucose) and the solution saturated with 95% O₂, 5% CO₂. The epithelial sodium channel was blocked with 100 μ M amiloride (Sigma-Aldrich). CFTR was stimulated by addition of forskolin (10 μ M) and genistein (50 μ M) to the apical side of the chamber followed by CFTR Inhibitor 172 (20 μ M, EMD Biosciences, apical) to isolate the CFTR-specific, apical Cl[–] current. Measurements were performed at 37°C and the short-circuit current (I_{sc}) was recorded and analysed with Acquire and Analyze 2.0 (Physiological Instruments).

CoPIT co-IP and sample preparation for multidimensional chromatography (LC/LC)–MS/MS. The detailed CoPIT protocol is available on the Nature protocol exchange website⁴⁴.

Rat monoclonal anti-CFTR antibody (3G11) was coupled to Protein G Sepharose 4 Fast Flow beads (GE Healthcare) at 6 mg ml^{–1} packed beads and covalently crosslinked to the beads with 20 mM dimethylpylimidate (DMP, Pierce). CFBE41o– or HBE41o– cells from passages 5 to 19 were grown to confluence in Advanced MEM supplemented with 10% FCS, 1% penicillin/streptomycin, 2 mM L-glutamine and additional appropriate antibiotics. Approximately 4×10^7 or $\sim 1 \times 10^8$ cells were harvested per IP, rinsed with PBS, lysed, CFTR protein complexes immunoprecipitated and prepared for MS analysis according to the CoPIT protocol. Briefly, cells were lysed on ice in TNI-buffer (0.5% Igepal CA-630 (Sigma-Aldrich), 50 mM Tris pH 7.5, 250 mM NaCl, 1 mM EDTA and 1 \times Complete EDTA-free Protease Inhibitor mix (Roche)). After water-bath sonication, insoluble material was removed by centrifugation (30 min, 18,000g, 4°C) and the supernatant pre-cleared by incubation with Sepharose CL-4B (GE Healthcare). The pre-cleared lysate was then incubated overnight at 4°C with 50 μ l (approximately 250 μ g) of anti-CFTR 3G11 antibody covalently coupled to Protein G Sepharose. Immunoprecipitates were recovered by centrifugation (500g, 5 min, 4°C), washed three times with lysis buffer and twice with lysis buffer containing no detergent. Bound proteins were eluted twice with 0.2 M glycine pH 2.3 and 0.5% Igepal CA-630 (30 min, 37°C) and precipitated (eluate:methanol:chloroform, 1:4:1, v:v:v). The precipitate was washed with 95% methanol and re-solubilized in 100 mM Tris pH 8.5 and 0.2% Rapigest (Waters). Samples were reduced with 5 mM TCEP (Pierce), alkylated with 10 mM iodoacetamide (Pierce) and proteins digested overnight with 3 μ g of sequencing-grade recombinant trypsin (Promega). Formic acid (9% final, v-v) was added to inactivate Rapigest (2 h, 37°C), any precipitate removed by centrifugation (15 min, 18,000g at room temperature), and samples reduced to near dryness *in vacuo*. To identify non-specific contaminating proteins, control IPs were performed from (1) isogenic CFTR null cells to identify background that is recognized by the 3G11 antibody and (2) by using mock-IPs, in which no antibody is coupled to the beads, to identify bead- and cell-specific background.

Expression profiling. Protein lysates from CFBE41o– and HBE41o– cells at the same passage number were prepared in TNI lysis buffer, precipitated (lysate:methanol:chloroform (1:4:1, v:v:v) and 100 μ g of protein reduced, alkylated and digested with trypsin as described above. Resulting peptides were labelled with 6-plex Tandem Mass Tag (TMT) labelling reagent (Thermo-Fisher) according to the manufacturer's recommendations. Subsequently, Rapigest was inactivated by acidification with 10% formic acid, insoluble precipitate removed by centrifugation (15 min, 18,000g) and samples reduced to near dryness *in vacuo*.

LC/LC–MS/MS. Samples were analysed by nano-electrospray ionization (ESI)–LC/LC–MS/MS on an LTQ–Orbitrap XL, LTQ or Orbitrap Elite (Thermo Fisher) by placing the triphasic MudPIT column in-line with an Agilent 1100 quaternary HPLC pump (Agilent) and separating the peptides in multiple dimensions with a modified six-step gradient containing 0%, 20%, 40%, 60%, and 100% of buffer C (500 mM ammonium acetate/5% acetonitrile/0.1% formic acid) over 12 h with the last step (100%) repeated, or a ten-step gradient (0%, 10%, 20%, 30%, 40%, 50%, 70%, 80%, 90%, 100% buffer C) over 20 h as described previously⁴⁵.

Each full-scan mass spectrum (400–2000 m/z) was followed by 6 (LTQ, LTQ-Orbitrap XL), or 20 (Orbitrap Elite) data-dependent MS/MS scans at 35% normalized collisional energy and an ion count threshold of 1,000 (LTQ-Orbitrap XL, Orbitrap Elite) or 500 counts (LTQ). Dynamic exclusion was used with an exclusion list of 500, repeat time of 60 s and asymmetric exclusion window of -0.51 and $+1.5$ Da. To avoid cross-contaminations between the different samples, each sample was loaded onto a fresh column.

CoPIT data analysis. Raw files were extracted with RawExtract (fields.scripps.edu/researchtools.php) and MS/MS spectra searched with ProLuCID⁴⁶ against the human International Protein Index database version 3.23, using a target–decoy approach in which each protein sequence was reversed and concatenated to the normal database⁴⁷. Search parameters were set to no enzyme specificity, 50 p.p.m. precursor mass tolerance and carboxamidomethylation ($m = 57.021464$ Da) as a static modification. Search results were filtered with DTASelect version 2.1 (ref. 48), allowing for tryptic peptides only and a peptide false discovery rate of less than 0.5%, usually corresponding to a protein false discovery rate of less than 1.0%. To uniformly control the false discovery rate across samples in CoPIT, and to facilitate comparison, sqt files of replicate samples were filtered in a single DTASelect run and split again in corresponding replicate subsets for further analysis. Samples with non-sufficient recovery of the bait (<35 spectral counts) were excluded from further analysis. To remove redundancy due to isoform-specific identifiers, which is problematic for statistical analysis, International Protein Index numbers were first converted to Entrez Gene symbols using the X-REF Converter developed by RIKEN (<http://refdic.rci.riken.jp/tools/xrefconv.cgi>) and manual annotation based on the Ensembl release 43 (<http://www.regulatorygenomics.org>); the highest PSM (peptide-spectrum match) values of all protein variants per gene and experiment were retained. CoPIT assumes that proteins binding non-specifically and non-selectively to carrier or antibody are detected with equal likelihood in experimental conditions (e) and control (c), as shown previously^{49,50}. Ratios for proteins p were calculated as $r_p = \log_{10}(\sum_{i=1}^n \text{PSM}_{p\ e\ i} / \sum_{i=1}^n \text{PSM}_{p\ c\ i})$, where n is the number of experiments. Data were then plotted in Matlab, and a bimodal model was applied to analyse the frequency distribution of all ratios $r_{p\ ec}$ and fitted with a Gaussian of two

terms $v_{\text{fit}} = A_{\text{bg}} \exp\left\{-\frac{1}{2}[(\rho - \mu_{\text{bg}})/\sigma_{\text{bg}}]^2\right\} + A_{\text{sp}} \exp\left\{-\frac{1}{2}[(\rho - \mu_{\text{sp}})/\sigma_{\text{sp}}]^2\right\}$ with a goodness of fit between $0.90 \leq R^2 \leq 0.98$, where (bg) is background- and (sp) bait-specific interactors. Confidence values were calculated for each protein according to $P = \text{erf}\left[|r_p - \mu_{\text{bg}}| / \sqrt{2 \times (\sigma_{p\ ec}^2 + \sigma_{\text{bg}}^2 + \sigma_{\text{sp}}^2)}\right]$ where σ_{bg} , μ_{bg} , σ_{sp} and

μ_{sp} were derived from the respective terms of the Gaussian fit. Proteins that were identified only in background control samples were eliminated from the analysis as obvious background contaminants. For a protein to be considered a potentially true interactor, we required further that it be detected in at least two independent biological replicates of the same condition to minimize random sampling errors and identities. Fold change of a protein p between two different experimental conditions was calculated according to

$$r_p = \log_{10} \left(\frac{\sum_{i=1}^n \text{PSM}_{p\ e\ i}}{\sum_{i=1}^n \text{PSM}_{p\ c\ i}} \right) / \left(\frac{\sum_{i=1}^n \text{PSM}_{\text{bait}\ e\ i}}{\sum_{i=1}^n \text{PSM}_{\text{bait}\ c\ i}} \right)$$

Sample sizes were not pre-determined with statistical methods in this discovery-based proteomic approach. Errors for relative changes were calculated on the basis of random error of measurement

$$\sigma_p = \frac{1}{2} \left(\log_{10} \left(1 + \sqrt{\frac{1}{\sum_{i=1}^n \text{PSM}_{p\ e\ i}} + \frac{1}{\sum_{i=1}^n \text{PSM}_{p\ c\ i}}} \right) + \log_{10} \left(1 - \sqrt{\frac{1}{\sum_{i=1}^n \text{PSM}_{p\ e\ i}} + \frac{1}{\sum_{i=1}^n \text{PSM}_{p\ c\ i}}} \right) \right)$$

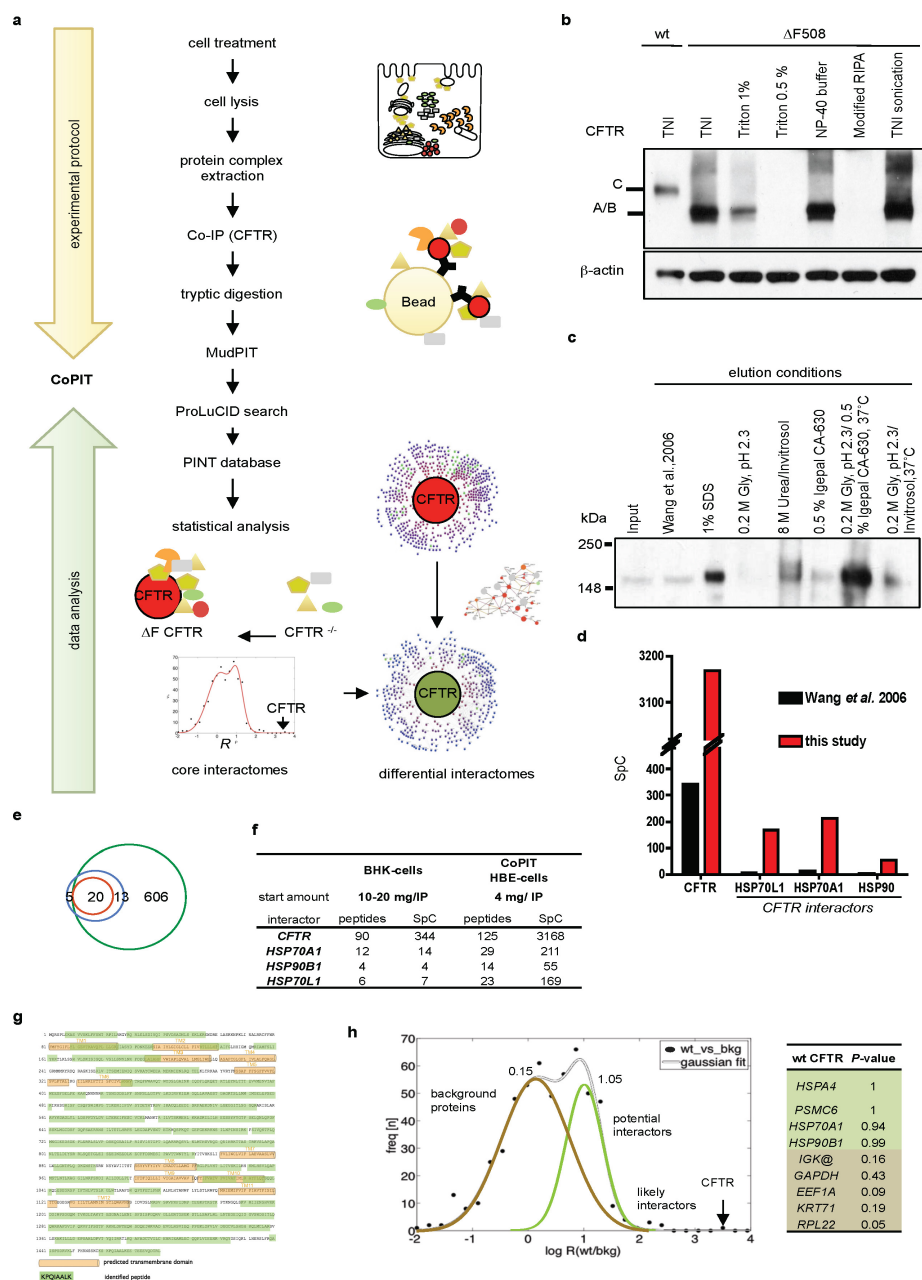
in CoPIT. If not indicated otherwise, the following significance definitions were used throughout all figures: *, $(\sigma_p + \log_{10} 1.32) \leq r_p < (2\sigma_p + \log_{10} 1.32)$; **, $r_p \geq (2\sigma_p + \log_{10} 1.32)$; wherein r_p is the average relative ratio of the protein and σ_p is the random error of measurement.

Annotation data were derived from Uniprot Knowledge Base, Entrez Gene information, GO Miner and literature review on PubMed. Interactions between the identified interactors were obtained with the GeneMANIA 2.2 Plugin⁵¹ in Cytoscape 2.8.2 using physical interactions reported in BIOGRID-small scale studies, BIOGRID and BIND as well as Pathway information reported in Pathway Commons. Proteins, their connections and according functional annotation were then graphed in Radial Topology Viewer 0.6, which was based on Medusa⁵², whereby length of individual edges reflects a quantitative relationship with the bait such as enrichment over background.

Analysis of additional small networks was performed using Osprey 1.2.0 (ref. 53) and Ingenuity Pathway Analysis (Ingenuity). Analysis of the expression profiling experiments was performed in Censur and the Integrated Proteomics pipeline IP2 (Integrated Proteomics Applications) using the TMT option with a tolerance of 10 millidaltons and a minimum intensity threshold of 100,000 relative ion counts⁵⁴. Statistical significance was determined with an unpaired t -test for differential expression (two-tailed and two-sample t -test on every protein). The volcano plot was generated with the biostatistics package in Matlab (Mathworks). The data set was uploaded to Proteomics INTEgrator (PINT; S.M.-B., unpublished observations) for online accession at <http://sealion.scripps.edu/pint?project=CFTR> ('CFTR' data set). It includes all qualitative and quantitative data over all experimental conditions and replicates measured. In addition, PINT provides an advanced query and annotation system, including the retrieval of Uniprot annotations assigned to the proteins in the data set.

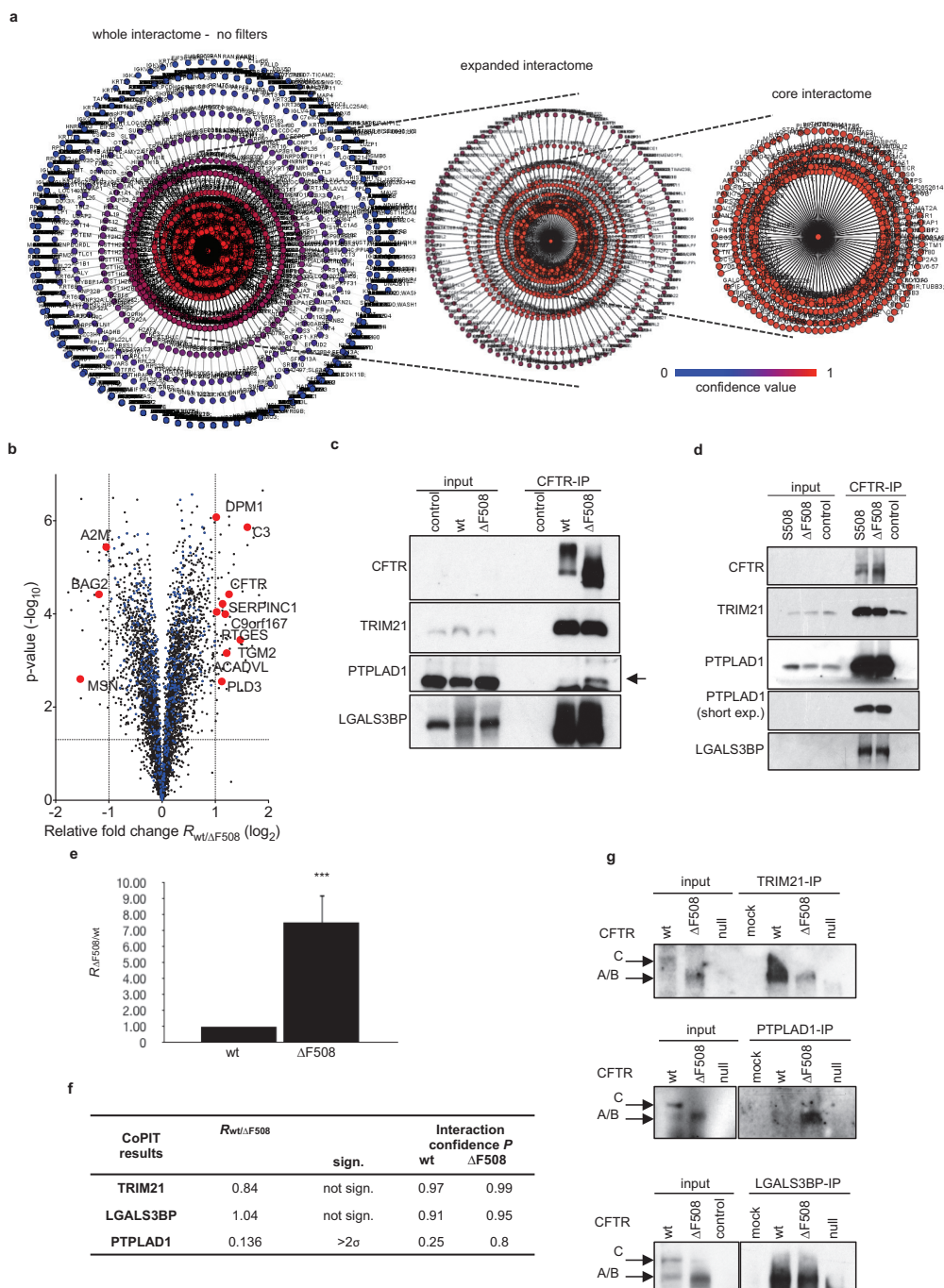
CFTR core interactome hierarchical clustering analysis. The CFTR interaction profile of a given condition was represented by \log_{10} -transformed ratios of core interactome protein abundances (sum of spectral counts across the replicates of that condition) and the abundance value of CFTR in that same condition. Hierarchical clustering of the different conditions was produced using the average linkage algorithm. The distance between two conditions was set to one minus their Pearson correlation. Heat-map representation was produced using gplots version 2.14.1, and bootstrap values were obtained using the R package pvclust 1.2-2 (ref. 55).

1. Tiscornia, G., Singer, O. & Verma, I. M. Production and purification of lentiviral vectors. *Nature Protocols* **1**, 241–245 (2006).
2. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).
3. Galletta, L. V., Jayaraman, S. & Verkman, A. S. Cell-based assay for high-throughput quantitative screening of CFTR chloride transport agonists. *Am. J. Physiol. Cell Physiol.* **281**, C1734–C1742 (2001).
4. Pankow, S., Bamberger, C., Calzolari, D., Bamberger, A. & Yates, J. R. Characterization of membrane protein interactomes by Co-interacting Protein Identification Technology (CoPIT). *Protocol Exchange* (submitted).
5. Washburn, M. P., Wolters, D. & Yates, J. R. III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnol.* **19**, 242–247 (2001).
6. Xu, T. et al. ProLuCID, a fast and sensitive tandem mass spectra-based protein identification program. *Mol. Cell Proteom.* **5**, S174 (2006).
7. Elias, J. E., Haas, W., Faherty, B. K. & Gygi, S. P. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nature Methods* **2**, 667–675 (2005).
8. Cociorva, D., Tabb, D. L. & Yates, J. R. in *Current Protocols in Bioinformatics* (ed. Bateman, A. et al.) Ch. 13, Unit 13.4 (Wiley, 2007).
9. Li, X. J., Zhang, H., Ranish, J. A. & Aebersold, R. Automated statistical analysis of protein abundance ratios from data generated by stable-isotope dilution and tandem mass spectrometry. *Analyt. Chem.* **75**, 6648–6657 (2003).
10. Ranish, J. A. et al. Identification of TFB5, a new component of general transcription and DNA repair factor IIH. *Nature Genet.* **36**, 707–713 (2004).
11. Montojo, J. et al. GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. *Bioinformatics* **26**, 2927–2928 (2010).
12. Pavlopoulos, G. A., Hooper, S. D., Sifrim, A., Schneider, R. & Aerts, J. Medusa: a tool for exploring and clustering biological networks. *BMC Res. Notes* **4**, 384 (2011).
13. Breitkreutz, B. J., Stark, C. & Tyers, M. Osprey: a network visualization system. *Genome Biol.* **4**, R22 (2003).
14. Park, S. K., Venable, J. D., Xu, T. & Yates, J. R., 3rd. A quantitative analysis software tool for mass spectrometry-based proteomics. *Nature Methods* **5**, 319–322 (2008).
15. Suzuki, R. & Shimodaira, H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006).



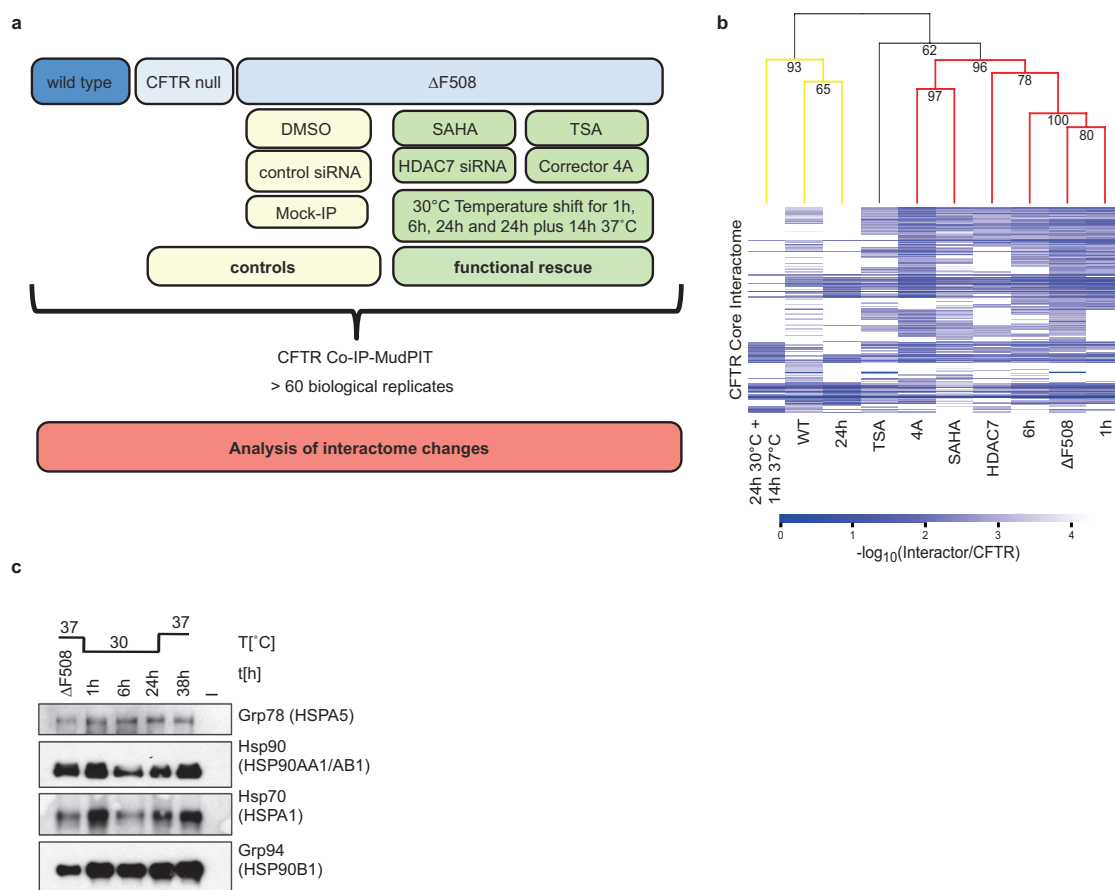
Extended Data Figure 1 | CoPIT workflow and results. **a**, Schematic overview of the Co-PIT workflow. Top: cell lysates for IP were prepared from $\geq 4 \times 10^7$ lung epithelial cells (CFBE41o⁻ or HBE41o⁻) with emphasis on extracting both cytoplasmic and membrane protein interactors of CFTR, pre-cleared before co-IP with anti-CFTR antibody 3G11. Proteins eluted from the beads were purified by methanol/chloroform precipitation and digested with trypsin, before loading onto a MudPIT column and online MudPIT data acquisition. Bottom: resulting spectra were searched with ProLuCID and search results filtered with DTASelect 2.1 to a protein false positive rate of $< 1\%$ before normalization and further statistical analysis of the data set. Core CFTR interactomes were determined by modelling the distribution functions of control and sample IPs, and applying corresponding confidence scores and abundance filters. Corresponding networks were graphed using Radial Topology Viewer and differential comparison performed. Data are stored in the PINT tool. **b**, Improved recovery of CFTR and interactors. Western blot depicting improved recovery of $\Delta F508$ CFTR from CFBE41o⁻ cells with TNI buffer compared with different lysis buffers. A, B and C indicate the different CFTR glycoforms. **c**, Western blot showing enhanced recovery of $\Delta F508$ CFTR from beads after co-IP with detergent and heat aided low pH elution compared with other directly MS-compatible elution methods. Lane entitled 'Wang *et al.* 2006': elution conditions as described in ref. 11. Gly, glycine. **d**, Enhanced sensitivity of the CFTR co-IP and chromatography is

reflected by enhanced spectral counts for CFTR itself and well-established interactors such as HSP70 and HSP90. **e**, Comparison between the CFTR interactome reported in ref. 11 and this study (Supplementary Table 4). Thirty-three of the reported 38 interactions in Calu-3 cells were recovered; 20 were confirmed as highly confident interactions (innermost circle) and 13 as medium confident interactions in this study, achieving an almost complete overlap of the two data sets. **f**, Table showing the recovery of CFTR and exemplary, well-characterized interactors in co-IPs of WT CFTR (BHK cells (from ref. 11) or HBE41o⁻ cells (this study)). **g**, Sequence coverage of the CFTR protein with MS. Green background indicates identified amino acids whereas orange highlights putative transmembrane (TM) domains of CFTR numbered from 1 to 12. **h**, Frequency distribution $N_{p\ ec}$ of all $r_{p\ ec}$ determined for the experimental condition WT CFTR to control condition. Individual points (black dots) indicate the individual $v_{p\ ec}$ values. The two-term Gaussian fit is shown in grey. The individual Gaussian describing the distribution of non-specific binding is coloured in brown, whereas the Gaussian describing the enrichment for weak specific interactors is indicated in light green. The black arrow marks the $r_{p\ ec}$ determined for CFTR, the bait protein. Right: example P values for well-known CFTR interactors (light green) and proteins commonly identified as background in co-IP experiments (light brown). Threshold for a high-confidence $\Delta F508$ -CFTR interactor was calculated at ≥ 0.92 .



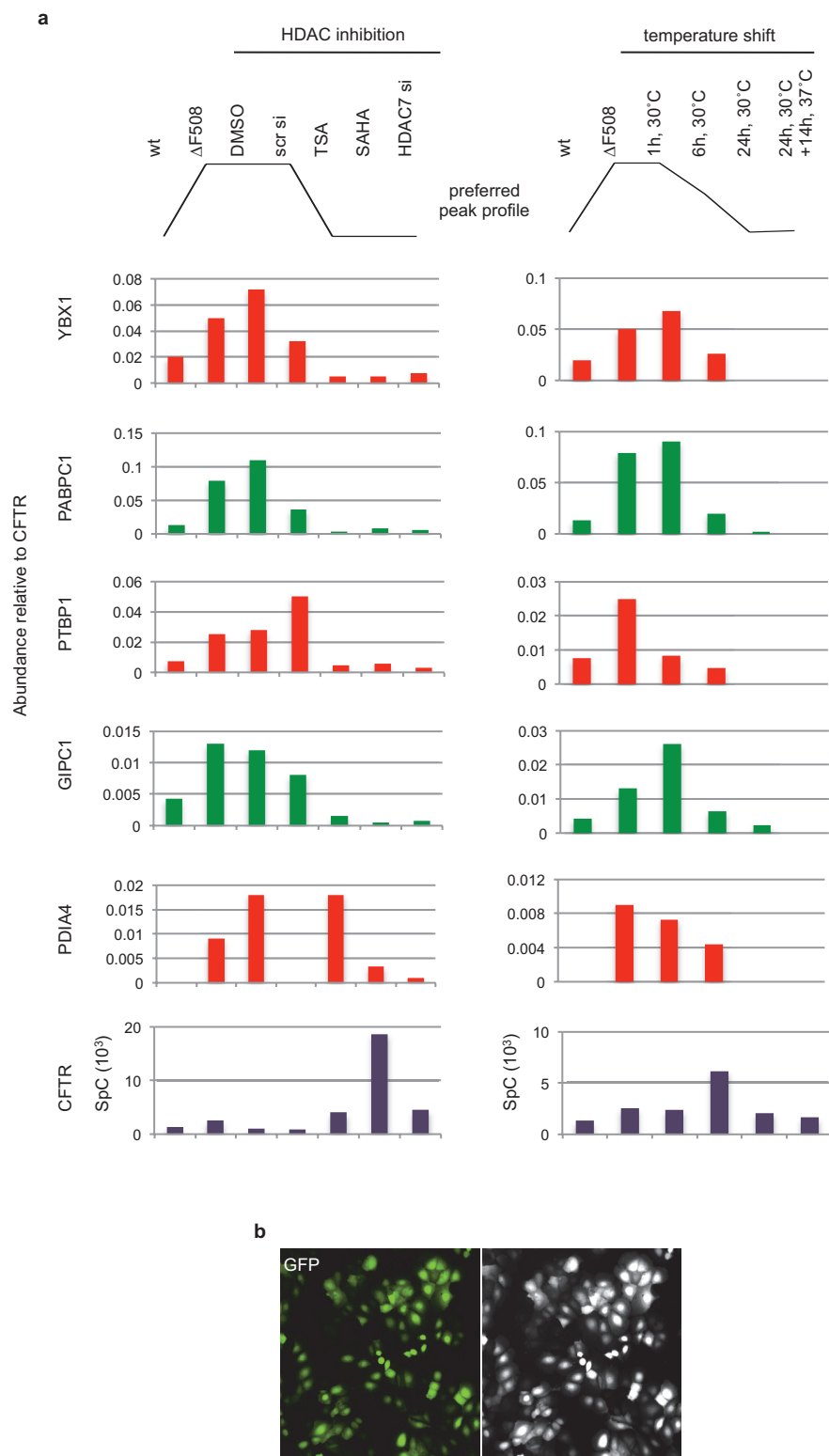
Extended Data Figure 2 | CFTR interactome and validation of novel interactors. **a**, Network representation of the $\Delta F508$ CFTR interactome in a radial topography map. The colour and relative distance to CFTR in the centre reflect the confidence P of an identified protein to be a specific CFTR interactor. Left: no filters were applied and all recovered proteins from the IPs are depicted. Right: core interactome of $\Delta F508$ CFTR ($P > 0.92$). Distance and colour indicate the confidence of an identified protein to be a specific CFTR interactor. **b**, Overlay of the interactome data with protein expression profiling data shows that observed interactome differences between WT and $\Delta F508$ CFTR are unrelated to expression changes between HBE410– and CFBE410– cells. The volcano plot displays the fold change and $\log_{10}(P)$ for 4,563 proteins quantified with tandem mass tag (TMT) in the expression profiling experiment. Core interactors of CFTR (529 proteins) that were not differentially regulated between the two cell lines are displayed in blue whereas significantly altered core interactors ($r \geq$ two-fold, $P < 0.01$) are displayed in red.

c, Western blotting of CFTR IPs confirms specific interaction of CFTR with the novel interaction partners TRIM21, LGALS3BP and PTPLAD1 in CFBE410– or HBE410– cells. Results indicate similar binding of WT and $\Delta F508$ CFTR with TRIM21 and LGALS3BP, and confirm enhanced binding of PTPLAD1 with $\Delta F508$ CFTR. **d**, CFTR co-IPs confirm CFTR interaction with TRIM21, PTPLAD1 and LGALS3BP in primary lung epithelial cells carrying either the $\Delta F508$ or the F508S mutation from a patient with CF. Control: CFTR null CFBE410– cell line. **e**, Ubiquitin (UBB/UBC) recovery is increased in $\Delta F508$ CFTR co-IPs. Error bars indicate mean \pm s.d. **f**, CoPIT confidence scores and observed fold changes for TRIM21, LGALS3BP and PTPLAD1 match recovery in the IP western blot. **g**, Reciprocal co-IP using newly identified, endogenous interactors as bait confirms interaction of TRIM21, LGALS3BP and PTPLAD1 with $\Delta F508$ CFTR and confirms differential binding of PTPLAD1 to WT and $\Delta F508$ CFTR. Control, null: CFTR null CFBE410– cell line; mock: beads only—IP with no antibody added.

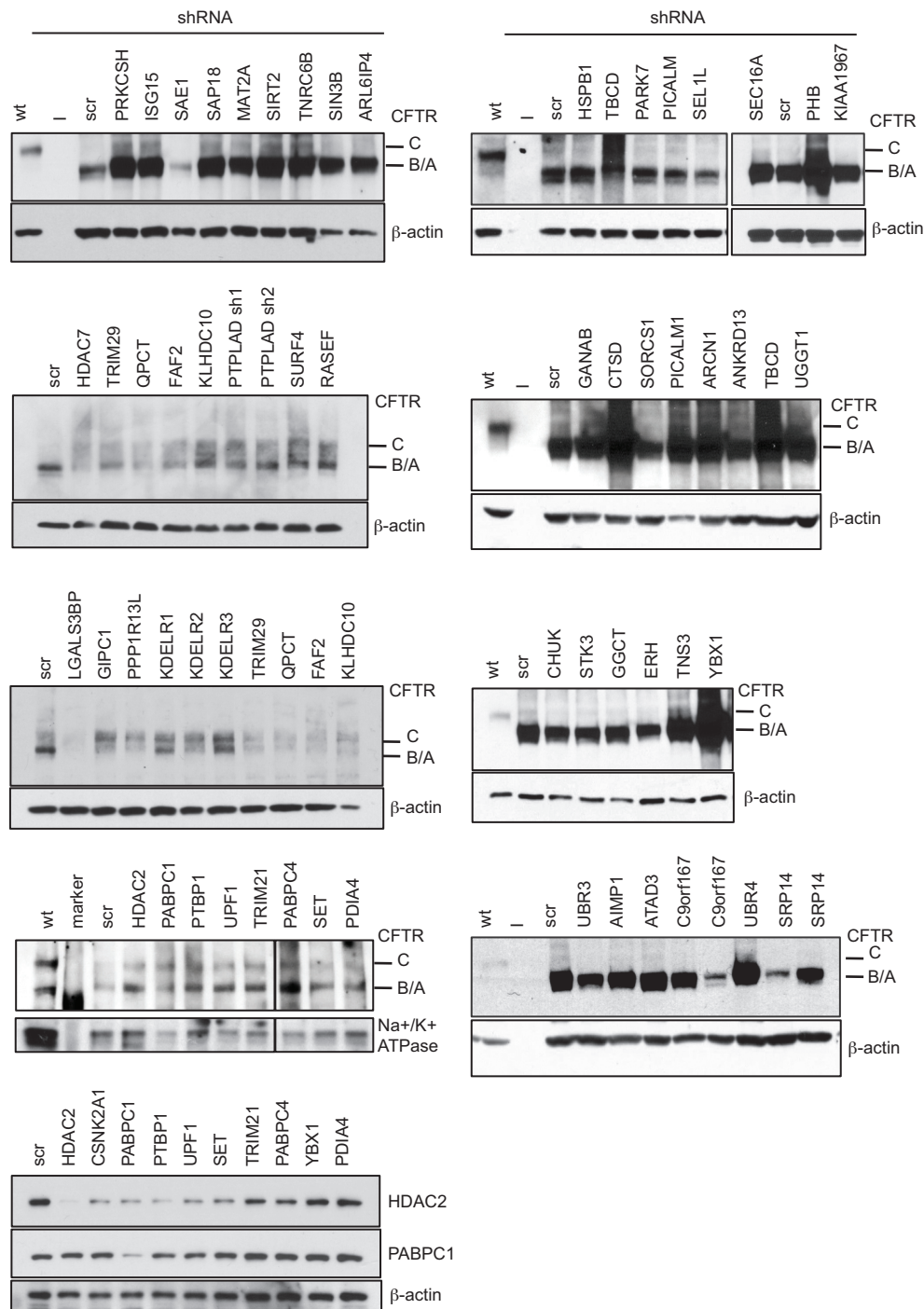


Extended Data Figure 3 | Overview of drug treatment, siRNA-mediated knockdown and temperature shift experiments. a, Schematic showing the experimental outline. **b**, Hierarchical clustering analysis of the CFTR core interactomes shows that the $\Delta F508$ CFTR interaction profile clusters with high significance with those of $\Delta F508$ CFTR at 1 h and 6 h temperature shifts to 30°C (mutant cluster), whereas temperature shift to 30°C for 24 h and temperature shift to 30°C for 24 h with reversal cause the respective $\Delta F508$ CFTR interaction profiles to significantly cluster with that of WT CFTR. Bootstrap values (10,000 samplings) are given for each tree node. Significant (bootstrap value > 90, yellow) and

highly significant clusters (bootstrap value > 95, red) are coloured on the dendrogram. The heat map indicates the relative protein abundance values measured by MS as negative \log_{10} ratios of interactors relative to CFTR. White in the heat map indicates that no interaction was observed. **c**, Expression of different heat-shock proteins. The western blot shows expression of HSP90 (encoded by HSP90AA1 and HSP90AB1), GRP78 (HSPA5), GRP94 (HSP90B1) and HSP70 (HSPA1) during temperature shift to 30°C. All data are from independent biological replicates, WT ($n = 7$), $\Delta F508$ ($n = 8$), SAHA ($n = 4$), TSA ($n = 4$), HDAC7 ($n = 3$), Cmpd 4a ($n = 3$).



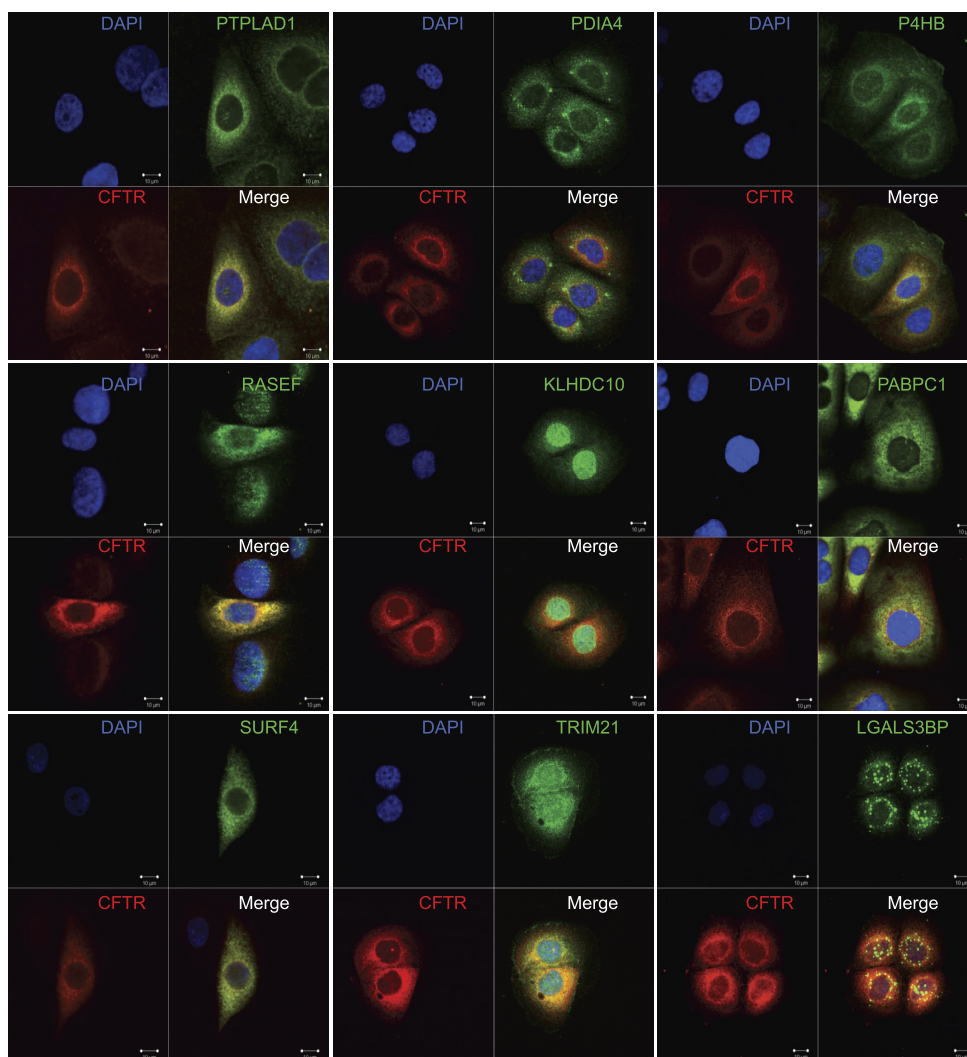
Extended Data Figure 4 | Interaction profiles of proteins selected for the RNAi screen. a, Observed interaction profiles of selected candidates and CFTR (bottom) and expected candidate profiles (top). **b,** Lentiviral infection rates were greater than 97% after 48 h in CFBE41o– cells as indicated by control green fluorescent protein (GFP) infection.



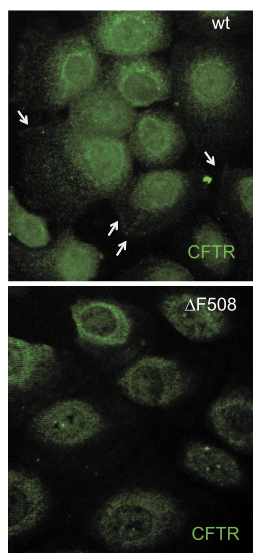
Extended Data Figure 5 | Western blot detection of Δ F508 CFTR upon RNAi of interactors. Δ F508 CFTR was detected 48–72 h after lentiviral shRNA infection using the 3G11 antibody or 24.1 antibody (lowest left panel). Rescue is indicated by appearance of band C. Detection of β -actin served as loading control. Samples on the same blot represent parallel

infections. Samples in the lower three left panels were lysed initially in TNF buffer, whereas samples in the other panels were lysed directly in $2\times$ Laemmli sample buffer as described in Methods. Scr, scrambled non-target shRNA.

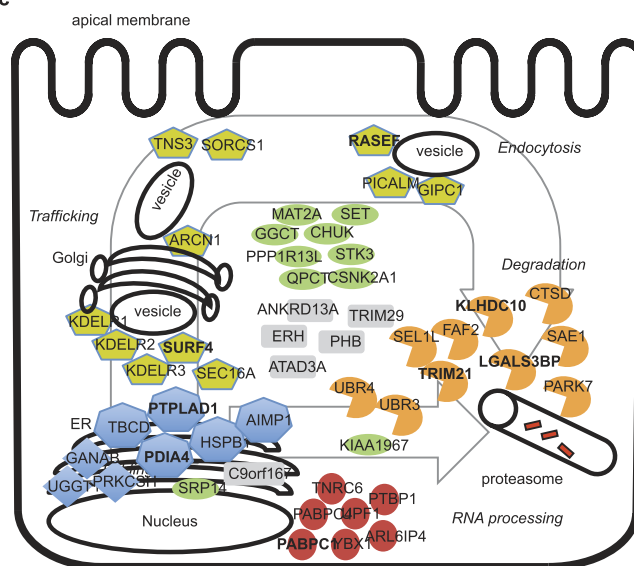
a



b

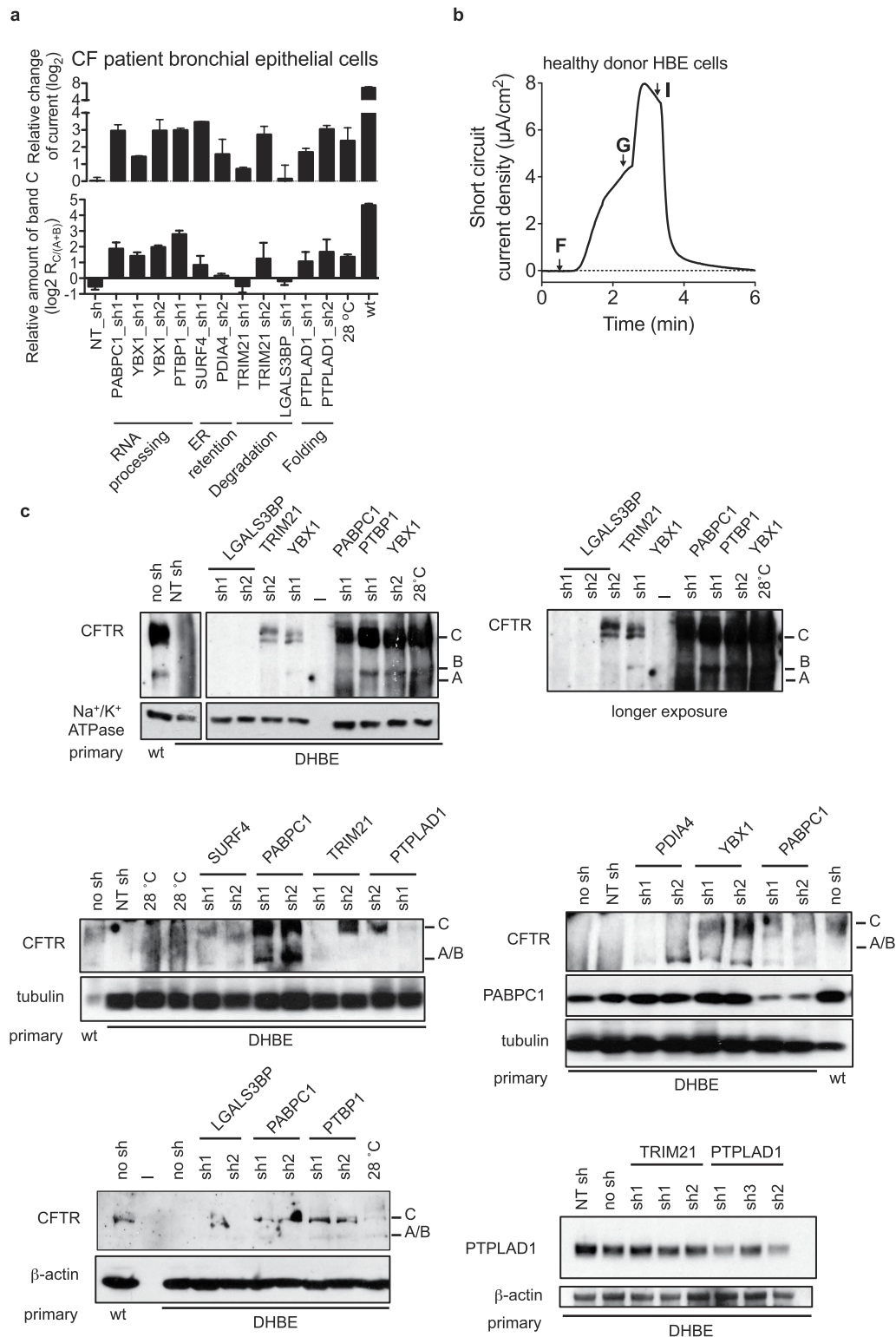


c



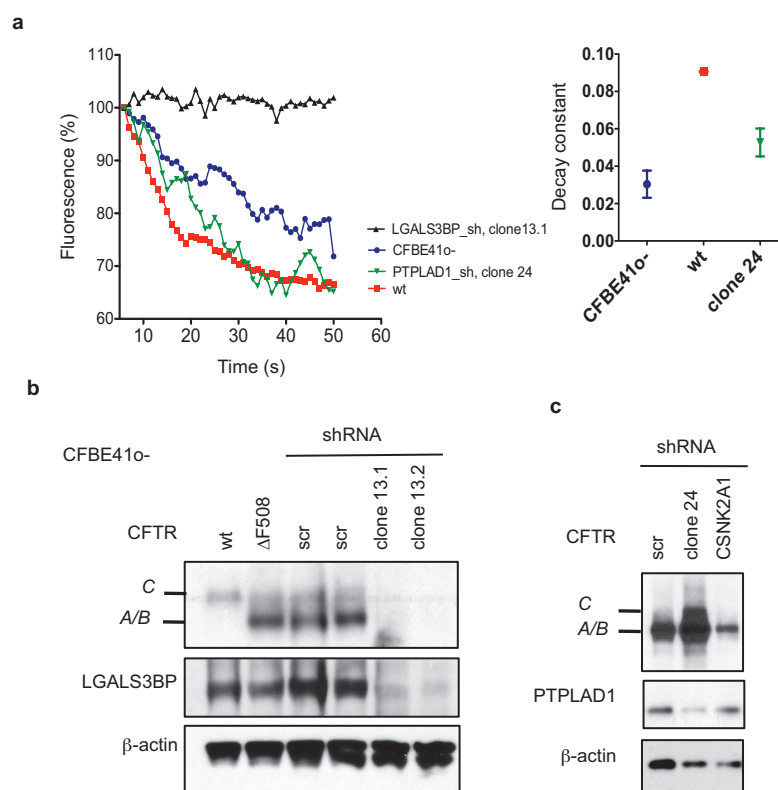
Extended Data Figure 6 | Co-localization of novel interactors with $\Delta F508$ CFTR. a, Each panel contains immunofluorescence staining of CFTR (red), interactor as indicated (green), nuclei (DAPI) and the merged picture. Scale bars, 10 μm . **b,** WT and $\Delta F508$ CFTR were detected by immunofluorescence staining (green) in HBE41o- and CFBE41o- cells, respectively. Arrows points to WT CFTR at the plasma membrane of

control cells. **c,** Schematic of a cell depicting sequential (spatio-temporal) regulation of $\Delta F508$ CFTR protein biogenesis by the interactors targeted in the shRNA screen. Functional classification of interactors is indicated by shape and colour. Proteins detected in co-localization studies are marked in bold.



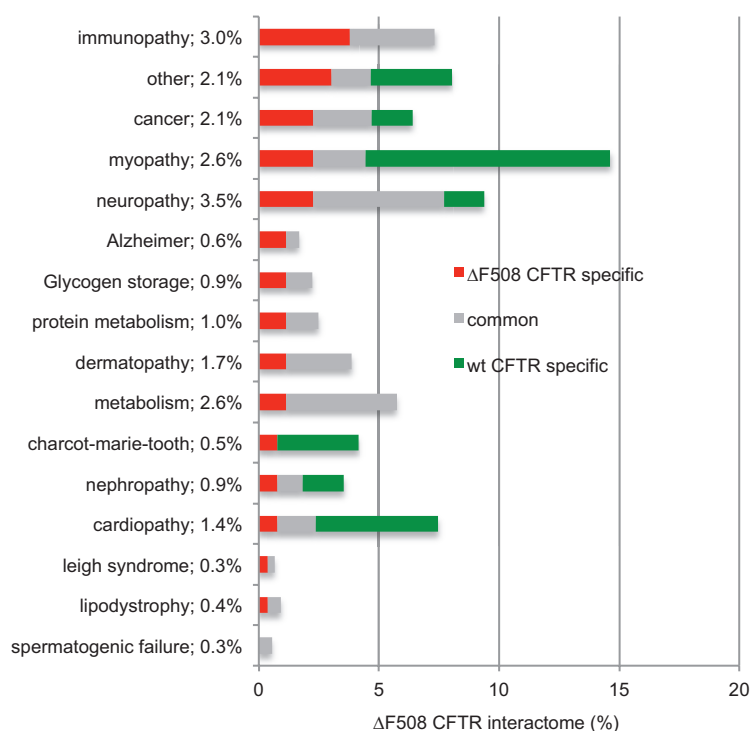
Extended Data Figure 7 | $\Delta F508$ CFTR detection in primary bronchial epithelial cells upon RNAi of key interactors. **a**, Quantification of the $\Delta F508$ CFTR ion channel activity (as fold change of the ΔI_{sc} relative to non-target shRNA) compared with the ratio of band C to band A/B in primary cells from a patient with CF or from a healthy donor (WT). Error bars indicate mean \pm s.e.m. **b**, Representative trace of forskolin ($10 \mu\text{M}$, F) and genistein ($50 \mu\text{M}$, G) activated, WT CFTR short-circuit current (I_{sc}) in a 30 d ALI culture from a healthy donor. CFTR inhibitor 172 (I) indicates

specificity of the measured I_{sc} for CFTR. **c**, Western blot of 28- to 30-day-old primary human bronchial epithelial Snapwell cultures from patients with CF (DHBE) indicates formation of band C after specific knockdown of PABPC1, YBX1, PTBP1, TRIM21, PTPLAD1 and SURF4 with different shRNAs. Tubulin, β -actin or Na⁺/K⁺-ATPase was used as a loading control. Knockdown of PABPC1 and PTPLAD1 was verified by western blotting with the respective antibodies. NT sh, non-target shRNA.



Extended Data Figure 8 | Halide assay results for CFTR chloride channel activity in stable cell clones. **a**, CFTR chloride channel activity was measured in HBE41o-, CFBE41o- and CFBE41o- cells with stable knockdown of LGALS3BP (clone 13) or PTPLAD1 (clone 24). Activity was measured by sodium-iodide-mediated quenching of a halide-sensitive Venus YFP. Time-lapse experiments show the iodide influx after pre-incubation of cells with 50 μ M genistein. Additional stimulation with forskolin was performed 15 s after addition of sodium iodide. Representative single cell traces are shown. Inset shows the fitted constant

for fluorescence decay time for each trace. **b**, Western blot showing the negative influence of LGALS3BP knockdown on Δ F508 CFTR protein stability. Clones 13.1 and 13.2 are two independent CFBE41o- clones that stably express an shRNA against LGALS3BP. The knockdown was validated by detection of LGALS3BP. **c**, Western blot showing increased production of Δ F508 CFTR band C in CFBE41o- cell clone 24 stably expressing an shRNA against PTPLAD1. The knockdown was validated by detection of PTPLAD1. Detection of β -actin served as loading control. Scr, scrambled non-target shRNA.



Extended Data Figure 9 | Percentage of CFTR interactors associated with known protein misfolding and other prevalent diseases. Bar graph showing the fraction of the interactome associated with genetic diseases listed in OMIM. Percentages next to the disease name indicate the percentage of Δ F508 CFTR-specific interactors involved in these diseases.

Interactors causative of Alzheimer disease and other neurodegenerative diseases such as Leigh syndrome are enriched in the Δ F508 CFTR interactome. 'Other' indicates diseases not fitting into one of the other categories listed.

DDX5 and its associated lncRNA *Rmrp* modulate T_H17 cell effector functions

Wendy Huang¹, Benjamin Thomas², Ryan A. Flynn³, Samuel J. Gavzy¹, Lin Wu¹, Sangwon V. Kim¹, Jason A. Hall¹, Emily R. Miraldi^{1,4,5,6}, Charles P. Ng¹, Frank W. Rigo⁷, Sarah Meadows⁸, Nina R. Montoya¹, Natalia G. Herrera¹, Ana I. Domingos⁹, Fraydoon Rastinejad¹⁰, Richard M. Myers⁸, Frances V. Fuller-Pace¹¹, Richard Bonneau^{4,5,6}, Howard Y. Chang³, Oreste Acuto² & Dan R. Littman^{1,12}

T helper 17 (T_H17) lymphocytes protect mucosal barriers from infections, but also contribute to multiple chronic inflammatory diseases. Their differentiation is controlled by ROR γ t, a ligand-regulated nuclear receptor. Here we identify the RNA helicase DEAD-box protein 5 (DDX5) as a ROR γ t partner that coordinates transcription of selective T_H17 genes, and is required for T_H17-mediated inflammatory pathologies. Surprisingly, the ability of DDX5 to interact with ROR γ t and coactivate its targets depends on intrinsic RNA helicase activity and binding of a conserved nuclear long noncoding RNA (lncRNA), *Rmrp*, which is mutated in patients with cartilage-hair hypoplasia. A targeted *Rmrp* gene mutation in mice, corresponding to a gene mutation in cartilage-hair hypoplasia patients, altered lncRNA chromatin occupancy, and reduced the DDX5-ROR γ t interaction and ROR γ t target gene transcription. Elucidation of the link between *Rmrp* and the DDX5-ROR γ t complex reveals a role for RNA helicases and lncRNAs in tissue-specific transcriptional regulation, and provides new opportunities for therapeutic intervention in T_H17-dependent diseases.

T_H17 cells are CD4⁺ lymphocytes that help to protect mucosal epithelial barriers against bacterial and fungal infections¹, and are also important in multiple autoimmune diseases^{2–7}. The T_H17 cell differentiation program is defined by the induced expression of ROR γ t (ref. 2), a sterol ligand-regulated nuclear receptor that focuses the activity of a cytokine-regulated transcriptional network upon a subset of key genomic target sites, including genes encoding the signature T_H17 cytokines (interleukin (IL)-17A, IL-17F and IL-22) as well as IL-23R, IL-1R1 and CCR6 (ref. 8). In mouse models, attenuation of ROR γ t activity results in protection from experimental autoimmune encephalomyelitis (EAE), T-cell-transfer-mediated colitis and collagen-induced arthritis^{2–5}. Like other nuclear receptors, ROR γ t interaction with its ligands results in recruitment of coactivators at regulated genomic loci⁹. We identified two new ROR γ t partners in T_H17 cells, an RNA helicase and a lncRNA, which together associate with ROR γ t to confer target-locus-specific activity in enabling the T-cell effector program.

The RNA helicase DDX5 functions in multiple cellular processes¹⁰, including transcription and ribosome biogenesis^{11–17}, in both a helicase-activity-dependent and -independent manner. The lncRNA *Rmrp*, RNA component of mitochondria RNA processing endoribonuclease (also known as RNase MRP), is highly conserved between mouse and human and is essential for early mouse development¹⁸. *Rmrp* was first identified as a component of the RNase MRP complex that cleaves mitochondrial RNAs¹⁹. In yeast, the *RMRP1* gene contributes to ribosomal RNA processing and regulates messenger RNA (mRNA) degradation²⁰. In humans, mutations located in evolutionarily conserved nucleotides at the promoter or within the transcribed region of *RMRP* result in cartilage-hair hypoplasia (CHH), a rare autosomal recessive disorder characterized by early childhood onset of

skeletal dysplasia, hypoplastic hair, defective immunity, predisposition to lymphoma and neuronal dysplasia of the intestine^{21,22}. Immune deficiency in CHH patients is associated with recurrent infections, haematological abnormalities and autoimmune pathologies in the joints and kidneys²³. The precise mechanisms by which *Rmrp* functions in the immune system have yet to be determined. Here we show that the helicase activity of DDX5 mediates *Rmrp*-dependent binding to ROR γ t and recruitment to a subset of chromatin target sites, thus controlling the differentiation of T_H17 cells at steady state and in animal models of autoimmunity.

DDX5 regulation of ROR γ t target genes

To identify novel interacting partners of ROR γ t in T_H17 cells, we enriched for endogenous ROR γ t-containing protein complexes and subsequently determined protein composition using liquid-chromatography–tandem mass spectroscopy (LC–MS/MS) (workflow shown in Extended Data Fig. 1a). Among the top hits of ROR γ t-interacting proteins was the RNA helicase DDX5. We validated this interaction through conventional co-immunoprecipitation experiments followed by immunoblot analysis (Extended Data Fig. 1b).

We investigated the function of DDX5 in T cells by breeding *Ddx5* conditional mutant mice with CD4-Cre mice to generate T-cell-specific DDX5-deficient mice (*Ddx5*^{fl/fl} CD4-Cre mice, denoted DDX5-T). DDX5-T mice were born at the expected Mendelian ratio, were fertile, and did not display any gross phenotypic abnormalities. The activation status of T cells in the periphery was similar between *Ddx5*^{+/+} CD4-Cre⁺ (wild type) and mutant mice (Extended Data Fig. 1c) that had no DDX5 protein in spleen and lymph node CD4⁺ T cells (Extended Data Fig. 1d). Naive CD4⁺ T cells sorted from wild-type and DDX5-T mice did not display differences in polarization towards

¹The Kimmel Center for Biology and Medicine of the Skirball Institute, New York University School of Medicine, New York, New York 10016, USA. ²Sir William Dunn School of Pathology, University of Oxford, Oxford OX1 3RE, UK. ³Center for Personal Dynamic Regulomes, Stanford University, Stanford, California 94305, USA. ⁴Center for Genomics and Systems Biology, Department of Biology, New York University, New York, New York 10003, USA. ⁵Courant Institute of Mathematical Sciences, Computer Science Department, New York University, New York, New York 10012, USA. ⁶Simons Center for Data Analysis, Simons Foundation, New York, New York 10010, USA. ⁷Isis Pharmaceuticals, Carlsbad, California 92010, USA. ⁸HudsonAlpha Institute for Biotechnology, Huntsville, Alabama 35806, USA. ⁹Instituto Gulbenkian de Ciencia, Oeiras 2780-156, Portugal. ¹⁰Integrative Metabolism Program, Sanford Burnham Prebys Medical Discovery Institute, Orlando, Florida 32827, USA. ¹¹Division of Cancer Research, University of Dundee, Dundee DD1 9SY, UK. ¹²Howard Hughes Medical Institute, New York University School of Medicine, New York, New York 10016, USA.

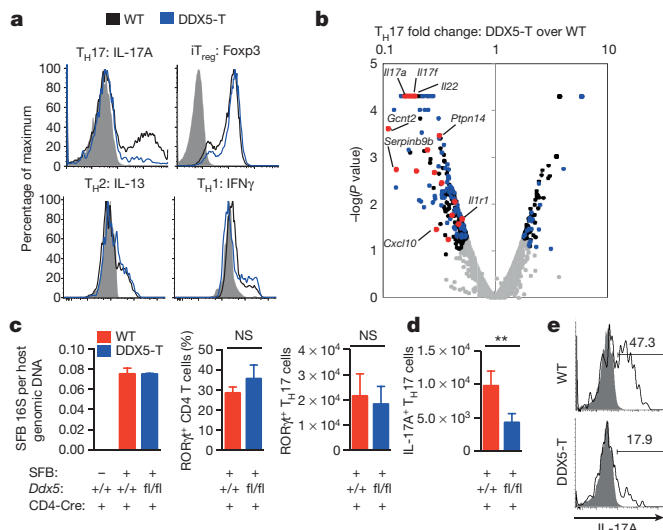


Figure 1 | Requirement for DDX5 in TH17 cytokine production *in vitro* and at steady state *in vivo*. **a**, Selective TH17 cell differentiation defect in DDX5-deficient T cells (DDX5-T) after polarization for 96 h. Representative of three independent experiments. WT, wild type. **b**, Volcano plot of RNA-seq of cultured TH17 cells from DDX5-T mice and littermate controls. Black dots, differentially expressed genes (minimum fold change of two with $P < 0.05$). Blue dots, known RORγt-dependent genes. Red dots, top RORγt-DDX5-coregulated genes. **c**, **d**, SFB colonization and percentage and number of RORγt⁺CD4⁺ T cells (**c**) and number of IL-17A-producing CD4⁺ T cells (**d**) in ileal lamina propria of co-housed wild-type (+/+; $n = 5$) and DDX5-T (fl/fl; $n = 5$) CD4-Cre⁺ mice. Graphs show mean \pm s.d. from two independent experiments, combined. NS, not significant. ** $P < 0.01$ (paired *t*-test). **e**, Representative IL-17A expression in CD4⁺Foxp3⁺RORγt⁺ TH17 cells from ileal lamina propria of wild-type and DDX5-T mice after restimulation.

TH1, TH2 and induced regulatory T (iTreg) cell phenotypes *in vitro* (Fig. 1a). In contrast, DDX5-T naive T cells cultured under TH17-polarizing conditions produced substantially less IL-17A than wild-type cells (Fig. 1a). RORγt protein expression and nuclear localization were similar between wild-type and DDX5-T TH17-polarized cells (Extended Data Fig. 1d, e), and, like RORγt, DDX5 protein localized mainly to the nucleus (Extended Data Fig. 1f). These results suggest that DDX5 is not required for TH17 lineage commitment, but contributes to TH17 cell effector functions.

DDX5 can function as a transcriptional coactivator^{12,24,25}, augmenting the activities of other nuclear receptor family members, including the oestrogen and androgen receptors^{12,26}. To determine whether DDX5 partners with RORγt to facilitate the TH17 cell transcriptional program, we performed RNA sequencing (RNA-seq) on *in vitro* polarized TH17 cells from wild-type or DDX5-T mice. Among the 325 genes that were significantly dysregulated in DDX5-deficient T cells 96 h after polarization, approximately 40% had been previously identified as RORγt targets in TH17 cells⁸ (Extended Data Fig. 2a). Ingenuity Pathway Analysis (Qiagen) of DDX5-RORγt-coregulated genes revealed enrichment in 'T helper cell differentiation program' as well as 'interleukin production' (Extended Data Fig. 2b). Coregulated genes (Fig. 1b) included those for the signature TH17 cytokines (*Il17a*, *Il17f* and *Il22*) (Extended Data Fig. 2c). Independent biological samples were used to validate a subset of RORγt target genes with and without altered expression in DDX5-deficient TH17 cells (Extended Data Fig. 2d).

We used anti-DDX5 antibodies in genome-wide chromatin immunoprecipitation sequencing (ChIP-seq) studies to identify DDX5-occupied loci. A specific subset of previously published RORγt-occupied loci, including *Il17a* and *Il17f*, were enriched for DDX5 co-localization, as determined by seqMINER clustering analysis

(Extended Data Fig. 3a, b). ChIP with quantitative PCR (ChIP-qPCR) was used to validate DDX5 enrichment at the *Il17a* and *Il17f* loci and its dependency on RORγt in polarized TH17 cells (Extended Data Fig. 3c). These results suggest that DDX5 overlaps with RORγt in modulating a specific subset of the TH17 cell transcriptional program.

DDX5 function *in vivo* in TH17 cells

At steady state, cytokine-producing TH17 cells populate the small intestinal lamina propria of animals colonized with commensal segmented filamentous bacteria (SFB)²⁷. When colonized with SFB, DDX5-T mice and wild-type littermates had similar numbers of ileal-residing Foxp3⁺RORγt⁺CD4⁺ TH17 cells (Fig. 1c). However, the number and proportion of IL-17A-producing cells among RORγt⁺CD4⁺ cells from DDX5-T mice were markedly reduced compared to wild-type littermate controls (Fig. 1d, e).

To evaluate the role of DDX5 in TH17-driven inflammation, we used a T-cell transfer model of colitis, in which disease severity is dependent on RORγt expression in donor T cells^{3,28}. Following transfer of CD4⁺CD45RB^{hi} T cells into Rag-deficient (*Rag2*^{-/-}) recipients, mice that received wild-type T cells experienced weight loss (Fig. 2a) and developed colitis (Fig. 2b), whereas recipients of DDX5-T cells did not. Total RNA from large intestine lamina propria mononuclear cells revealed a significant reduction of both *Il17a* and *Ifng* transcripts from recipients of DDX5-T cells compared to wild-type controls (Extended Data Fig. 4a). Interestingly, there were comparable proportions of IFNγ-producing CD4⁺RORγt⁺T-bet⁺ (conventional TH1) cells in recipients of T cells from either wild-type or DDX5-T mice (Extended Data Fig. 4b). However, recipients of cells from DDX5-T mice displayed a significant reduction in CD4⁺Foxp3⁺RORγt⁺ T cells co-expressing IL-17A and IFNγ, an important feature of pathogenic T cells in several inflammatory disease settings^{2,29,30} (Fig. 2c and

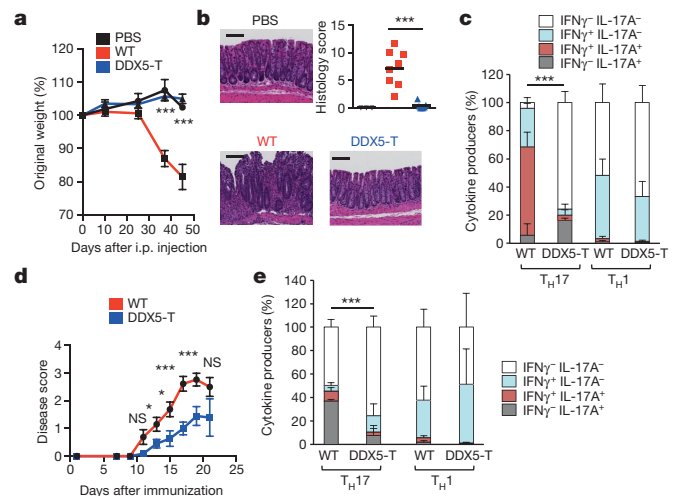


Figure 2 | Role of DDX5 in mouse models of TH17-cell-mediated autoimmune disease. **a**, Weight change in *Rag2*^{-/-} recipients of wild-type or DDX5-T CD4⁺ naive T cells in the transfer model of colitis measured on days 0, 10, 25, 37 and 45 (PBS, $n = 4$; wild type, $n = 9$; DDX5-T, $n = 13$, combined from three independent experiments). i.p., intraperitoneal. **b**, Haematoxylin and eosin (H&E) staining and analysis of large intestine at day 45. Representative sections (scale bars, 100 μm) and histology scores (scale of 0–24) are shown. Scores for PBS ($n = 3$), wild-type (red, $n = 8$) and DDX5-T (blue, $n = 7$) mice are from two independent experiments. **c**, Cytokine production defect in DDX5-T TH17 (RORγt⁺) but not TH1 (RORγt⁺T-bet⁺) cells in large intestine lamina propria at day 45 ($n = 4$ per group). **d**, EAE disease scores (scale of 0–5) in co-housed myelin oligodendrocyte glycoprotein (MOG)-immunized littermates. Wild-type ($n = 13$) and DDX5-T ($n = 11$) mice, combined from three independent experiments. **e**, Defective IL-17A production in DDX5-T CD4⁺RORγt⁺ cells in the spinal cord of MOG-immunized mice ($n = 7$ per group). Graphs show mean \pm s.d. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (unpaired, *t*-test).

Extended Data Fig. 4b). Consistent with a loss of pathogenic capacity, DDX5-T mice also exhibited attenuated disease compared to wild-type controls during EAE (Fig. 2d). Analysis of spinal cord infiltrates after immunization revealed a reduced proportion of IL-17A-producing CD4⁺ T cells (Fig. 2e and Extended Data Fig. 4c). Consistent with our *in vitro* findings, these results in mice indicate that DDX5 selectively regulates the T_H17 effector program, both in steady state and under inflammatory conditions.

Function of DDX5-associated lncRNA

RNA helicases are highly conserved enzymes that utilize the energy derived from ATP hydrolysis to unwind RNA duplexes, facilitate RNA annealing, and displace proteins from RNA. It was previously shown that DDX5 transcriptional coactivator activity for oestrogen receptor, androgen receptor and the transcription factor RUNX2 is independent of RNA helicase activity^{12,24,26}. We tested this requirement in the context of ROR γ t by retroviral transduction of DDX5-deficient T cells cultured under T_H17-polarizing conditions with expression constructs for wild-type or mutant DDX5 with an inactivated helicase domain (helicase-dead). Surprisingly, only wild-type DDX5 rescued IL-17A and IL-17F production in these polarized T_H17 cells (Fig. 3a, b and Extended Data Fig. 5a). This result suggested that perhaps RNA substrate(s) for the helicase activity of DDX5 contribute to its transcriptional coactivator role in T_H17 cells.

We next searched for RNA molecules that might participate in DDX5-ROR γ t-mediated transcription in T_H17 cells. We first depleted ribosome-bound mRNAs undergoing active protein synthesis. Lysates pre-cleared of ribosomes were then subjected to RNA immunoprecipitation (RIP) with antibodies specific for DDX5 or ROR γ t, followed by deep sequencing of the associated RNAs (RIP-seq). Among 49,893 annotated lncRNAs in the mouse RefSeq and NONCODE database, 2,533 noncoding RNAs were expressed in T_H17 cells (fragments per kilobase of transcripts per million mapped reads (FPKM) > 1, Extended Data Fig. 5b). Interestingly, the steroid receptor RNA activator (SRA) lncRNA, previously found to be associated with DDX5 in muscle cells¹⁵, was not enriched in DDX5-containing protein complexes in T_H17 cells. Instead, we found *Rmrp* to be the most enriched RNA associated with DDX5 and, to a lesser degree, ROR γ t, in T_H17 cells (Fig. 3c and Extended Data Fig. 5c). RIP-qPCR with independent biological samples confirmed enrichment of *Rmrp* RNA in DDX5 pull-downs from T_H17 cells, but not from thymocyte lysates (Extended Data Fig. 5d).

RNA fluorescence *in situ* hybridization revealed that *Rmrp* is localized in the nucleus of T_H17 cells (Extended Data Fig. 6a). To evaluate the functional role of *Rmrp*, we transiently depleted *Rmrp* RNA from primary mouse T_H17 cells using an RNaseH-dependent antisense oligonucleotide (ASO). Similar to the DDX5-deficient T_H17 cells, cells depleted of *Rmrp* expressed reduced *Il17a* and *Il17f* mRNA (Fig. 3d and Extended Data Fig. 6b). Human T_H17 cells also displayed reduced cytokine production upon depletion of RMRP or DDX5 (Fig. 3e and Extended Data Fig. 6c), suggesting that this regulatory mechanism is evolutionarily conserved. Notably, *Rmrp* RNA knockdown in DDX5-deficient mouse T_H17 cells did not further reduce IL-17A and IL-17F expression (Fig. 4a). Expression of ROR γ t-dependent, but DDX5-independent, CCR6 was unaffected by the reduction in *Rmrp*. Transduction of *Rmrp* into T cells cultured in T_H1-polarization conditions had little effect on IFN γ production, but there was marked enhancement of IL-17A and IL-17F production in wild-type, but not DDX5-deficient, cells cultured in T_H17-polarization conditions (Fig. 4b and Extended Data Fig. 7a, b). Thus, *Rmrp*-dependent cytokine gene expression requires the presence of DDX5.

T_H17 program in *Rmrp* mutant mice

In contrast to wild-type *Rmrp*, a mutant *Rmrp* carrying a single nucleotide change (270G > T), corresponding to an allele identified in CHH patients (262G > T), failed to potentiate IL-17A production

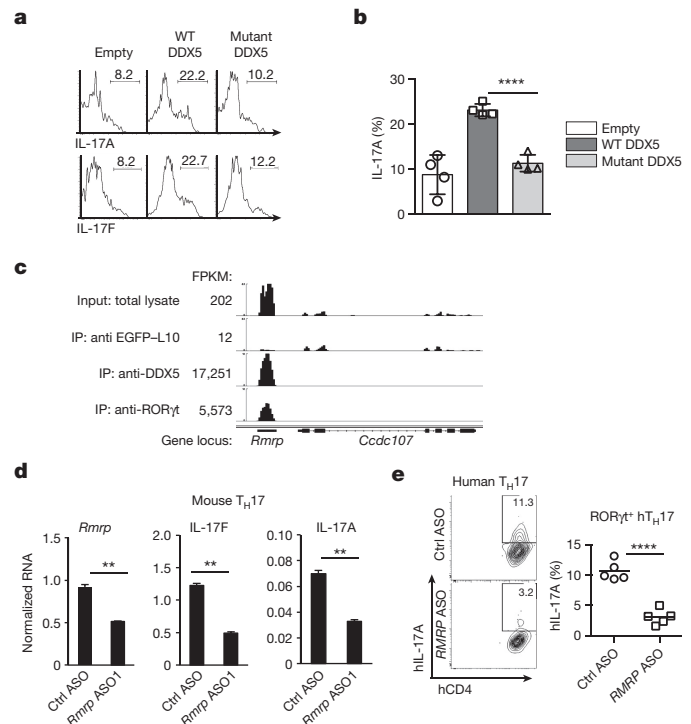


Figure 3 | Requirement for helicase-competent DDX5 and its associated lncRNA *Rmrp* in induction of T_H17 cell cytokines. **a, b**, Cytokine production in DDX5-T cells transduced with wild-type or helicase-mutant DDX5 and subjected to sub-optimal T_H17 cell polarization. Results from four independent experiments shown. **c**, IGV browser view of *Rmrp* showing coverage of mapped RNA reads from total T_H17 lysate, ribosome TRAP-seq (EGFP-L10; described in Methods), DDX5 RIP-seq and ROR γ t RIP-seq. IP, immunoprecipitate. **d**, Effect of mouse *Rmrp*-specific ASO. Results are representative of three independent experiments with two technical replicates. Ctrl, control. **e**, IL-17A production following RMRP knockdown in *in vitro* polarized human T_H17 cells. Each data point (right panel) represents cells from a healthy donor (*n* = 5). Graphs show mean \pm s.d. ***P* < 0.01; *****P* < 0.0001 (unpaired, *t*-test).

after transduction into T_H17-polarized cells (Extended Data Fig. 7c, d). To investigate whether G270 of *Rmrp* contributes to ROR γ t transcriptional output *in vivo*, we generated mice homozygous for the *Rmrp* G270T point mutation, using CRISPR-Cas9 (clustered, regularly interspaced short palindromic repeats coupled with CRISPR-associated proteins) technology (Fig. 4c). The mice were born at the expected Mendelian ratios and had no gross defects. ROR response element (RORE)-regulated luciferase activity was reduced in transiently transfected T_H17 cells from DDX5-deficient and *Rmrp*^{G270T} mice and after ASO-mediated knockdown of *Rmrp* (Fig. 4d). Comparison of the transcription profiles of *in vitro* polarized T_H17 cells from wild-type, ROR γ t-deficient, DDX5-deficient and *Rmrp*^{G270T/G270T} mice indicated that 96 ROR γ t-dependent T_H17 cell genes were coregulated by *Rmrp* together with DDX5 (Extended Data Fig. 7e and Fig. 4e). Reverse transcription (RT)-qPCR analysis of independent biological samples from *in vitro* polarized T cells from wild-type and *Rmrp*^{G270T/G270T} mice confirmed reduced *Il17f* mRNA expression in the latter (Extended Data Fig. 7f), despite a similar amount of ROR γ t binding to known *cis*-regulatory loci (Extended Data Fig. 7g). The proportion of ROR γ t⁺Foxp3⁻ T_H17 cells among total ileal lamina propria CD4-lineage cells was unaffected in *Rmrp*^{G270T/G270T} mice, but these cells expressed relatively little IL-17A compared to those in wild-type littermates (Fig. 4f). Transfer of *Rmrp*^{G270T/G270T} T cells into *Rag2*^{-/-} mice resulted in reduced colitis, as determined by weight loss and colon histology, compared to the transfer of wild-type cells (Extended Data Fig. 8a). These phenotypes are similar to those observed in recipients of DDX5-deficient T-cells

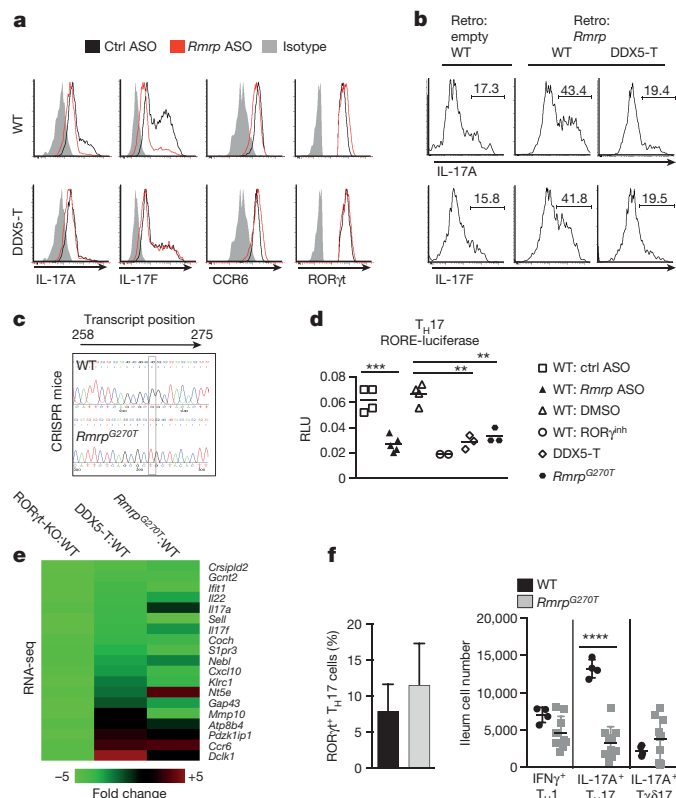


Figure 4 | Analysis of DDX5-dependent *Rmrp* function in T_H17 cell differentiation. **a, b,** Cytokines in wild-type and DDX5-T *in vitro* polarized T_H17 cells after *Rmrp* knockdown (**a**) or overexpression (**b**). Representative of three independent experiments. **c,** Sequence of the *Rmrp* gene (nucleotides 258–275) from wild-type and *Rmrp*^{G270T/G270T} littermates. **d,** *Rmrp*-dependent expression of a RORE-directed firefly luciferase reporter nucleofected into polarized T_H17 cells at 72 h. Firefly and control *Renilla* luciferase activities were measured 24 h later. Each dot represents the result from one nucleofection. Results from two independent experiments. DMSO, dimethyl sulfoxide; RLU, relative luciferase units; RORγ^{inh}, RORγ antagonist ML209. **e,** Top RORγt targets coregulated by DDX5 and *Rmrp*. **f,** Proportion of CD4⁺Foxp3[−] T cells expressing RORγt (left) and numbers of T_H1 (IFNγ⁺RORγ[−]Tbet⁺), T_H17 (IL-17A⁺RORγ⁺Foxp3[−]) and T_H17 (IL-17A⁺Tγδ⁺RORγ⁺) cells (right) in small intestine lamina propria. Symbols represent cells from one mouse. Graphs show mean ± s.d. ****P* < 0.01; *****P* < 0.0001 (unpaired, *t*-test).

(Fig. 2a–c), which is consistent with an important role of *Rmrp* G270 in executing the T_H17 effector program *in vivo*.

RORγt and its closely related isoform RORγ perform distinct functions in diverse tissues. RORγt is critical for thymocyte development, regulating the survival of double-positive CD4⁺CD8⁺ cells, and development of secondary and tertiary lymphoid organs mediated by lymphoid tissue inducer cells³¹. While DDX5 and *Rmrp* are ubiquitously expressed, *Rmrp* was less enriched in thymocyte-derived than in T_H17 -cell-derived DDX5 immunoprecipitates (Extended Data Fig. 5d). When *Ddx5* was inactivated at the common lymphoid progenitor stage, by breeding the conditional mutant mice with IL7R-Cre mice, there was no apparent defect in thymocyte subset development (Extended Data Fig. 8b). Similarly, *Rmrp*^{G270T} knock-in mice displayed normal thymocyte subsets and also had intact secondary lymphoid organ development (Extended Data Fig. 8c). Together, these results suggest that the DDX5–*Rmrp* complex performs T_H17 -specific functions.

Rmrp in DDX5–RORγt complex formation

We next investigated how *Rmrp* contributes to the DDX5–RORγt-regulated transcriptional circuit in T_H17 cells. RORγt–DDX5 complex

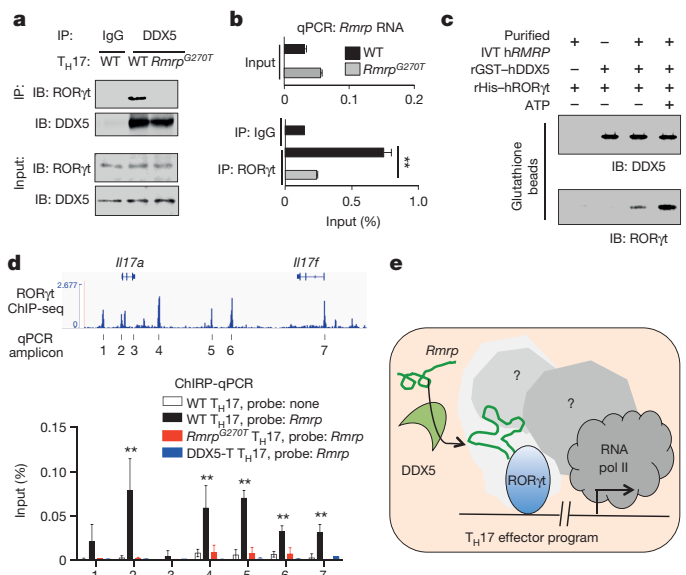


Figure 5 | *Rmrp* localization at RORγt-occupied genes and role in RORγt–DDX5 assembly. **a,** RORγt association with immunoprecipitated DDX5 in polarized T_H17 cells. IB, immunoblot. Representative of three independent experiments. **b,** *Rmrp* quantification by RT–qPCR in RORγt immunoprecipitates from polarized T_H17 cells. Representative of two independent experiments with two technical replicates. **c,** *Rmrp* requirement for ATP-dependent *in vitro* interaction of recombinant (r) glutathione S-transferase (GST)–DDX5 and His–RORγt. Representative of three independent experiments. IVT, *in vitro* transcribed RNA. For uncropped gels (**a, c**), see Supplementary Fig. 1. **d,** *Rmrp* occupancy at RORγt genomic target loci in polarized T_H17 cells. *Rmrp* ChIP–qPCR amplicons (bottom) are indicated in IGV browser view of RORγt ChIP–seq at the *Il17a* locus (top). Data from 2–4 experiments with two technical replicates. **e,** Model for DDX5–*Rmrp* complex recruitment to RORγt-occupied chromatin in T_H17 cells. Graphs show mean ± s.d. ***P* < 0.01 (unpaired, *t*-test).

assembly was severely compromised in T_H17 cells from *Rmrp*^{G270T} mice (Fig. 5a). Moreover, *Rmrp* recruitment to the RORγt protein complex was significantly reduced in T_H17 cells from *Rmrp* mutant animals (Fig. 5b). *In vitro*, *Rmrp* binds directly to recombinant DDX5 (Extended Data Fig. 9a). Notably, *Rmrp* was recruited to RORγt in the presence of wild-type, but not helicase-dead, DDX5. Furthermore, *in vitro* transcribed *Rmrp* RNA promoted RORγt interaction with wild-type, but not helicase-dead, DDX5 in the presence of ATP (Fig. 5c and Extended Data Fig. 9b). Mutant *Rmrp* was also defective in mediating DDX5–RORγt complex assembly *in vitro* (Extended Data Fig. 9c, d).

To determine whether *Rmrp* is associated with specific genomic loci, we performed chromatin isolation by RNA purification (ChIRP) followed by either locus-specific qPCR or deep sequencing (ChIRP–seq)³². We used two orthogonal antisense probe sets that specifically and robustly recovered *Rmrp* from T_H17 cells (Extended Data Fig. 10a). When combined for *Rmrp* ChIRP–qPCR, the probes recovered RORγt-bound regions in the *Il17a* and *Il17f* loci from T_H17 -polarized cells of wild-type but not DDX5-T or *Rmrp*^{G270T/G270T} mice, in an RNase-sensitive manner (Fig. 5d and Extended Data Fig. 10b). For ChIRP–seq, we focused our analysis on signals that overlapped with separate use of the two probe sets. HOMER motif analyses of *Rmrp* peak regions identified the ETS, DR2/RORE and AP1 transcription factor motifs to be the most highly enriched (Extended Data Fig. 10c). Consistent with this, *Rmrp* ChIRP–seq significantly overlapped with RORγt-bound loci, but not with sites occupied by CTCF or by other T_H17 transcription factors, such as BAF1, IRF4, STAT3 and c-Maf (Extended Data Fig. 10d). There was also significant overlap with RNA polymerase II (Pol II)- and histone H3 lysine 4 trimethylation (H3K4me3)-associated chromatin, which mark actively transcribed regions. Concordantly, ChIRP–seq of *Rmrp* in DDX5-T T_H17

cells revealed a loss of called *Rmrp* peaks (Extended Data Fig. 10e), consistent with a DDX5 contribution to *Rmrp* association with chromatin. *Rmrp* association with ROR γ t-bound sites was also reduced in polarized T_H17 cells from *Rmrp*^{G270T/G270T} mice despite a similar amount of RNA recovery (Extended Data Fig. 10f). Together, these results indicate that G270 of *Rmrp* is critical for DDX5–ROR γ t complex assembly and *Rmrp* recruitment to ROR γ t-occupied loci to coordinate the T_H17 effector program *in trans*.

Discussion

Nuclear lncRNAs have key roles in numerous biological processes³³ including adaptive and innate immunity^{34,35}, but how individual lncRNAs perform their activities and whether they contribute to immunological diseases remain unknown. We identified nuclear *Rmrp* as a key DDX5-associated RNA required to promote assembly and regulate the function of ROR γ t transcriptional complexes at a subset of critical genes implicated specifically in the T_H17 effector program (model in Fig. 5e). *Rmrp* thus acts *in trans* on multiple ROR γ t-dependent genes, and does so only upon interaction with enzymatically active DDX5 helicase. RNA-helicase-dependent functions of lncRNAs have been described, for example, the *Drosophila* male cell-specific lncRNAs roX1 and roX2 that are modified by the MLE helicase to enable expression of X-chromosome genes^{36,37}. In addition, DDX21 helicase activity in HEK293 cells is required for 7SK RNA regulation of polymerase pausing at ribosomal genes³⁸. Our results extend the concept of RNA helicase/lncRNA function to lineage-specific regulation of transcriptional programs.

Notably, unlike most lncRNAs, *Rmrp* is highly conserved among mammals. In humans, mutations of evolutionarily conserved nucleotides at the promoter or within the transcribed region of *RMRP* result in CHH^{21,22}. T cells from mice carrying a single nucleotide change (270G > T) in *Rmrp*, corresponding to one found in CHH patients (262G > T), had a compromised T_H17 cell effector program. CHH patients have been noted to have defective T-cell-dependent immunity, which may in part reflect reduced *RMRP*-dependent activity at ROR γ t target genes. As forced expression of either DDX5 or *Rmrp* enhanced T_H17 cytokine production, it is also possible that gain-of-function mutations in either of these molecules may contribute to T_H17-dependent inflammatory diseases.

ROR γ t is an attractive therapeutic target for multiple autoimmune diseases^{5,39}. However, ROR γ t and ROR γ have several other functions that would probably be affected by targeting of their shared ligand-binding pocket. ROR γ t is required for the development of early thymocytes, lymphoid tissue inducer cells that initiate lymphoid organogenesis³¹, type 3 innate lymphoid cells that produce IL-22 and protect epithelial barriers, and for IL-17 production by ‘innate-like’ T cells, including T cell receptor (TCR) $\gamma\delta$ and natural killer T cells^{40–43}. In the liver, ROR γ contributes to regulation of metabolic functions⁴⁴. Mechanisms by which ROR γ t and ROR γ differentially regulate transcription in these diverse cell types remain poorly understood. DDX5 and *Rmrp* are abundantly expressed in developing T cells in the thymus and in peripheral naive and differentiated T-helper subsets (W.H., unpublished observations). Notably, the contribution of DDX5–*Rmrp* to ROR γ t-dependent functions appears to be confined to T_H17 cells, as their loss of function did not affect thymocyte or lymphoid organ development. Our results raise the prospect that tissue- or cell-type-specific mechanisms exist to regulate how RNA helicases and their associated lncRNAs are assembled with distinct transcriptional complexes to promote diverse gene expression programs.

We speculate that the function of DDX5–*Rmrp* may be induced in response to specific tissue microenvironments *in vivo*. T_H17 cells differentiate at mucosal barriers in response to signals from the microbiota, and upregulate their expression of IL-17A locally^{45,46}. Regional signals may induce DDX5/*Rmrp* association with ROR γ t, resulting in the transcriptional activation of multiple loci that enable T_H17 cell effector functions. Our finding that DDX5 was required for the

differentiation of ‘pathogenic’ T_H17 cells^{2,29,30} suggests that strategies to interfere with this function may be of therapeutic benefit. A better understanding of this transcriptional regulatory system may provide new approaches for therapeutic intervention in autoimmune diseases and immune deficiencies in CHH patients.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 June; accepted 2 November 2015.

Published online 16 December 2015.

- Weaver, C. T., Hatton, R. D., Mangan, P. R. & Harrington, L. E. IL-17 family cytokines and the expanding diversity of effector T cell lineages. *Annu. Rev. Immunol.* **25**, 821–852 (2007).
- Ivanov, I. I. et al. The orphan nuclear receptor ROR γ t directs the differentiation program of proinflammatory IL-17⁺ T helper cells. *Cell* **126**, 1121–1133 (2006).
- Leppkes, M. et al. ROR γ -expressing Th17 cells induce murine chronic intestinal inflammation via redundant effects of IL-17A and IL-17F. *Gastroenterology* **136**, 257–267 (2009).
- Genovese, M. C. et al. LY2439821, a humanized anti-interleukin-17 monoclonal antibody, in the treatment of patients with rheumatoid arthritis: a phase I randomized, double-blind, placebo-controlled, proof-of-concept study. *Arthritis Rheum.* **62**, 929–939 (2010).
- Huh, J. R. et al. Digoxin and its derivatives suppress T_H17 cell differentiation by antagonizing ROR γ t activity. *Nature* **472**, 486–490 (2011).
- Kondo, Y. et al. Involvement of ROR γ t-overexpressing T cells in the development of autoimmune arthritis in mice. *Arthritis Res. Ther.* **17**, 105 (2015).
- Gaffen, S. L., Jain, R., Garg, A. V. & Cua, D. J. The IL-23–IL-17 immune axis: from mechanisms to therapeutic testing. *Nature Rev. Immunol.* **14**, 585–600 (2014).
- Ciofani, M. et al. A validated regulatory network for Th17 cell specification. *Cell* **151**, 289–303 (2012).
- O'Malley, B. W. & Kumar, R. Nuclear receptor coregulators in cancer biology. *Cancer Res.* **69**, 8217–8222 (2009).
- Huang, Y. & Liu, Z. R. The ATPase, RNA unwinding, and RNA binding activities of recombinant p68 RNA helicase. *J. Biol. Chem.* **277**, 12810–12815 (2002).
- Fuller-Pace, F. V. & Moore, H. C. RNA helicases p68 and p72: multifunctional proteins with important implications for cancer development. *Future Oncol.* **7**, 239–251 (2011).
- Clark, E. L. et al. The RNA helicase p68 is a novel androgen receptor coactivator involved in splicing and is overexpressed in prostate cancer. *Cancer Res.* **68**, 7938–7946 (2008).
- Linder, P. & Jankowsky, E. From unwinding to clamping — the DEAD box RNA helicase family. *Nature Rev. Mol. Cell Biol.* **12**, 505–516 (2011).
- Arun, G., Akhade, V. S., Donakonda, S. & Rao, M. R. mrhl RNA, a long noncoding RNA, negatively regulates Wnt signaling through its protein partner Ddx5/p68 in mouse spermatogonial cells. *Mol. Cell. Biol.* **32**, 3140–3152 (2012).
- Caretti, G. et al. The RNA helicases p68/p72 and the noncoding RNA SRA are coregulators of MyoD and skeletal muscle differentiation. *Dev. Cell* **11**, 547–560 (2006).
- Lin, C., Yang, L., Yang, J. J., Huang, Y. & Liu, Z. R. ATPase/helicase activities of p68 RNA helicase are required for pre-mRNA splicing but not for assembly of the spliceosome. *Mol. Cell. Biol.* **25**, 7484–7493 (2005).
- Jalal, C., Uhlmann-Schiffler, H. & Stahl, H. Redundant role of DEAD box proteins p68 (Ddx5) and p72/p82 (Ddx17) in ribosome biogenesis and cell proliferation. *Nucleic Acids Res.* **35**, 3590–3601 (2007).
- Rosenbluh, J. et al. *RMRP* is a non-coding RNA essential for early murine development. *PLoS ONE* **6**, e26270 (2011).
- Hsieh, C. L. et al. The gene for the RNA component of the mitochondrial RNA-processing endoribonuclease is located on human chromosome 9p and on mouse chromosome 4. *Genomics* **6**, 540–544 (1990).
- Esakova, O. & Krasilnikov, A. S. Of proteins and RNA: the RNase P/MRP family. *RNA* **16**, 1725–1747 (2010).
- Mäkitie, O., Kaitila, I. & Savilahti, E. Susceptibility to infections and *in vitro* immune functions in cartilage-hair hypoplasia. *Eur. J. Pediatr.* **157**, 816–820 (1998).
- Bonafé, L. et al. Evolutionary comparison provides evidence for pathogenicity of *RMRP* mutations. *PLoS Genet.* **1**, e47 (2005).
- Bacchetta, J. et al. Autoimmune hypoparathyroidism in a 12-year-old girl with McKusick cartilage hair hypoplasia. *Pediatr. Nephrol.* **24**, 2449–2453 (2009).
- Jensen, E. D. et al. p68 (Ddx5) interacts with Runx2 and regulates osteoblast differentiation. *J. Cell. Biochem.* **103**, 1438–1451 (2008).
- Dardenne, E. et al. RNA helicases DDX5 and DDX17 dynamically orchestrate transcription, miRNA, and splicing programs in cell differentiation. *Cell Rep.* **7**, 1900–1913 (2014).
- Wortham, N. C. et al. The DEAD-box protein p72 regulates ER α /oestrogen-dependent transcription and cell growth, and is associated with improved survival in ER α -positive breast cancer. *Oncogene* **28**, 4053–4064 (2009).
- Ivanov, I. I. et al. Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell* **139**, 485–498 (2009).

28. Powrie, F. *et al.* Inhibition of Th1 responses prevents inflammatory bowel disease in scid mice reconstituted with CD45RB^{hi} CD4⁺ T cells. *Immunity* **1**, 553–562 (1994).
29. Hirota, K. *et al.* Fate mapping of IL-17-producing T cells in inflammatory responses. *Nature Immunol.* **12**, 255–263 (2011).
30. Wang, Y. *et al.* The transcription factors T-bet and Runx are required for the ontogeny of pathogenic interferon- γ -producing T helper 17 cells. *Immunity* **40**, 355–366 (2014).
31. Sun, Z. *et al.* Requirement for ROR γ in thymocyte survival and lymphoid organ development. *Science* **288**, 2369–2373 (2000).
32. Chu, C., Quinn, J. & Chang, H. Y. Chromatin isolation by RNA purification (ChIRP). *J. Vis. Exp.* (2012).
33. Bonasio, R. & Shiekhattar, R. Regulation of transcription by long noncoding RNAs. *Annu. Rev. Genet.* **48**, 433–455 (2014).
34. Gomez, J. A. *et al.* The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon- γ locus. *Cell* **152**, 743–754 (2013).
35. Carpenter, S. *et al.* A long noncoding RNA mediates both activation and repression of immune response genes. *Science* **341**, 789–792 (2013).
36. Maenner, S., Muller, M., Frohlich, J., Langer, D. & Becker, P. B. ATP-dependent roX RNA remodeling by the helicase maleless enables specific association of MSL proteins. *Mol. Cell* **51**, 174–184 (2013).
37. Ilik, I. A. *et al.* Tandem stem-loops in roX RNAs act together to mediate X chromosome dosage compensation in *Drosophila*. *Mol. Cell* **51**, 156–173 (2013).
38. Calo, E. *et al.* RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. *Nature* **518**, 249–253 (2015).
39. Yang, J., Sundrud, M. S., Skepner, J. & Yamagata, T. Targeting Th17 cells in autoimmune diseases. *Trends Pharmacol. Sci.* **35**, 493–500 (2014).
40. Lee, Y. J., Holzapfel, K. L., Zhu, J., Jameson, S. C. & Hogquist, K. A. Steady-state production of IL-4 modulates immunity in mouse strains and is determined by lineage diversity of iNKT cells. *Nature Immunol.* **14**, 1146–1154 (2013).
41. Takatori, H. *et al.* Lymphoid tissue inducer-like cells are an innate source of IL-17 and IL-22. *J. Exp. Med.* **206**, 35–41 (2009).
42. Luci, C. *et al.* Influence of the transcription factor ROR γ t on the development of NKp46⁺ cell populations in gut and skin. *Nature Immunol.* **10**, 75–82 (2009).
43. Chien, Y. H., Zeng, X. & Prinz, I. The natural and the inducible: interleukin (IL)-17-producing $\gamma\delta$ T cells. *Trends Immunol.* **34**, 151–154 (2013).
44. Kang, H. S. *et al.* Gene expression profiling reveals a regulatory role for ROR α and ROR γ in phase I and phase II metabolism. *Physiol. Genomics* **31**, 281–294 (2007).
45. Yang, Y. *et al.* Focused specificity of intestinal T_H17 cells towards commensal bacterial antigens. *Nature* **510**, 152–156 (2014).
46. Sano, T. *et al.* An IL-23R/IL-22 circuit regulates epithelial serum amyloid A to promote local effector Th17 responses. *Cell* (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. V. Pokrovskii for unpublished ATAC-seq data and L. X. Garmire for suggestions on our manuscript. This work was supported by a Cancer Research Institute Irvington Postdoctoral Fellowship (W.H.), Institutional NRSA T32 CA009161_Levy (W.H.), National Multiple Sclerosis Society postdoctoral fellowship FG 2089-A-1 (L.W.), Career Development Award (329388) from the Crohn's and Colitis Foundation of America (S.V.K.), Dale and Betty Frey Fellowship of the Damon Runyon Cancer Research Foundation 2105-12 (J.A.H.), HHMI Exceptional Research Opportunities Program (N.R.M. and N.H.), NIH F30 1F30CA189514-01 (R.A.F.), NIH P50-HG007735 and R01HG004361 (H.Y.C.), NIH R01AI080885 (D.R.L.), NIH R01DK103358 (R.B. and D.R.L.), and the Howard Hughes Medical Institute (H.Y.C. and D.R.L.).

Author Contributions W.H. and D.R.L. designed experiments, analysed data and wrote the manuscript with input from the co-authors; B.T. and O.A. performed mass spectrometry studies; F.W.R. designed and synthesized control and *Rmrp* ASOs; S.J.G. and L.W. performed MOG-EAE immunization and blinded scoring; S.V.K. performed blinded histology scoring on colitis sections; W.H. and A.I.D. designed and performed ribosome TRAP-seq studies. S.M. and R.M.M. performed library preparation for RNA sequencing studies; N.R.M. and N.G.H. performed microscopy studies; F.R. provided recombinant full length His-tagged hROR γ t, and F.V.F.-P. generated DDX5 conditional mutant animals. J.A.H. performed ROR γ t ChIP studies. C.P.N. performed DDX5 studies in the thymus. R.A.F., W.H. and H.Y.C. performed ChIRP-seq experiments. E.R.M. and R.B. performed statistical analyses on ChIRP-seq experiments.

Author Information RNA-seq, TRAP-seq, RIP-seq, and ChIRP-seq data have been deposited in the Gene Expression Omnibus under accession number GSE70110. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.R.L. (dan.littman@med.nyu.edu).

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized. In vivo transfer colitis and EAE mouse experiments were blinded, but cell culture and in vitro studies were not.

Mice. EEF1A1-LSL-EGFP-L10 (lox-stop-lox-EGFP-L10 knockin at the *Efla1* locus) transgenic mice, ROR γ t-deficient animals and *Ddx5*^{fl/fl} mice have been previously described elsewhere^{31,47,48}. Conditional mutant mice were bred to CD4-Cre transgenic animals (Taconic) and maintained on the C57BL/6 background. We bred heterozygous mice to yield 6–8-week-old *Ddx5*^{+/+} CD4-Cre⁺ (subsequently referred to as wild type) and *Ddx5*^{fl/fl} CD4-Cre⁺ (referred to as DDX5-T) littermates for experiments examining DDX5 in peripheral T-cell function. DDX5 conditional mutant mice were also bred to IL7R-Cre transgenic animals (Jackson Laboratory) (with *Ddx5* deleted in common early lymphoid progenitors; referred to as DDX5-clpKO) for experiments examining DDX5 functions during T-cell development in the thymus. *Rmrp*^{G270T} knock-in mice were generated using CRISPR-Cas9 technology by the Rodent Genetic Engineering Core (RGEN) at New York University Langone Medical Center. Guide RNA and homology directed repair donor template sequences are provided in Supplementary Table 1. Heterozygous crosses provided *Rmrp*^{+/+} (wild-type) and *Rmrp*^{G270T/G270T} littermates for *in vivo* studies. All animal procedures were in accordance with protocols approved by the Institutional Animal Care and Use Committee of the New York University School of Medicine (Animal Welfare Assurance number: A3435-01).

In vivo studies. Steady-state small intestines were collected for isolation of lamina propria mononuclear cells as previously described⁴⁵. For detecting SFB colonization, SFB-specific 16S primers were used⁴⁹. Universal 16S and/or host genomic DNA was quantified simultaneously to normalize SFB colonization of each sample. All primer sequences are listed in Supplementary Table 1.

For the adoptive transfer model of colitis, 5×10^5 CD4⁺CD25[−]CD44^{low}CD45RB^{hi}CD62L^{hi} T cells were isolated from mouse splenocytes by FACS sorting and administered i.p. into *Rag2*^{−/−} mice as previously described⁵⁰. Animal weights were measured approximately weekly. Between weeks seven and eight, large intestines were collected for H&E staining and isolation of lamina propria mononuclear cells as previously described⁴⁵. The H&E slides from each sample were examined in a double-blind fashion. The histology scoring (scale 0–24) was based on the evaluation of criteria described previously⁵¹.

For induction of active EAE, each mouse was immunized subcutaneously on day 0 with 100 µg of MOG35–55 peptide, emulsified in CFA (Complete Freund's Adjuvant supplemented with additional 2 mg ml^{−1} *Mycobacterium tuberculosis*), and injected i.p. on days 0 and 2 with 100 ng per mouse of pertussis toxin (Calbiochem). The EAE scoring system was as follows: 0, no disease; 1, limp tail; 2, weak/partially paralysed hind legs; 3, completely paralysed hind legs; 4, complete hind and partial front leg paralysis; 5, complete paralysis/death.

In transfer colitis and EAE experiments, animals of different genotypes were co-housed and weighed and scored blindly. For statistical power level of 0.8, probability level of 0.05, anticipated effect size of 2, minimum sample size per group for two-tailed hypothesis is 6. Two-tailed unpaired Student's *t*-test was performed using Prism (GraphPad Software). We treated a *P* value of less than 0.05 as a significant difference. All experiments were performed at least twice.

In vitro T-cell culture and phenotypic analysis. Mouse T cells were purified from lymph nodes and spleens of 6–8-week-old mice, by sorting live (DAPI[−]), CD8[−]CD19[−]CD4⁺CD25[−]CD44^{low/int}CD62L⁺ naive T cells using a FACS Aria (BD). Detailed antibody information is provided in Supplementary Table 1. Cells were cultured in Iscove's Modified Dulbecco's Medium (IMDM, Sigma) supplemented with 10% heat-inactivated FBS (Hyclone), 50 U penicillin-streptomycin (Invitrogen), 4 mM glutamine and 50 µM β-mercaptoethanol. For T-cell polarization, 200 µl of cells was seeded at 0.3×10^5 cells per ml in 96-well plates pre-coated with goat anti-hamster IgG at a 1:20 dilution of stock (1 mg ml^{−1}, MP Biomedicals). Naive T cells were activated with anti-CD3 (2.5 µg ml^{−1}) and anti-CD28 (10 µg ml^{−1}). Cells were cultured for 4–5 days under T_H17-polarizing conditions (0.1–0.3 ng ml^{−1} TGF-β, 20 ng ml^{−1} IL-6), T_H1- (10 ng ml^{−1} IL-12, 10 U ml^{−1} IL-2), T_H2- (10 ng ml^{−1} IL-4, 10 U ml^{−1} IL-2), or T_{reg}- (5 ng ml^{−1} TGF-β, 10 U ml^{−1} IL-2) conditions.

Human T cells were isolated from peripheral blood of healthy donors using anti-human CD4 MACS beads (Miltenyi). Human CD4⁺ T cells were cultured in 96-well U bottom plates in 10 U ml^{−1} of IL-2, 10 ng ml^{−1} of IL-1β, 10 ng ml^{−1} of IL-23, 1 µg ml^{−1} of anti-IL-4, 1 µg ml^{−1} of anti-IFNγ and anti-CD3/CD28 activation beads (Life Technologies) at a ratio of 1 bead per cell, as previously described⁵².

For cytokine analysis, cells were incubated for 5 h with phorbol 12-myristate 13-acetate (PMA) (50 ng ml^{−1}; Sigma), ionomycin (500 ng ml^{−1}; Sigma) and GolgiStop (BD). Intracellular cytokine staining was performed according to the manufacturer's protocol (Cytofix/Cytoperm buffer set from BD Biosciences and FoxP3 staining buffer set from eBioscience). A LSR II flow cytometer (BD

Biosciences) and FlowJo (Tree Star) software were used for flow cytometry and analysis. Dead cells were excluded using the Live/Dead fixable aqua dead cell stain kit (Invitrogen).

Nucleic acid reagents and T-cell transduction. Custom *Rmrp* and predesigned *Malat1* Stellaris RNA fluorescence *in situ* hybridization (FISH) probes were purchased from BiosearchTech and used to label mouse *Rmrp* and *Malat1* RNA in cultured T_H17 cells according to the manufacturer's protocol. Control and human DDX5-specific short interfering RNAs (siRNAs) were obtained from Cell Signaling. Synthesis of ASOs was performed as previously described⁵³. All ASOs were 20 nucleotides in length and had a phosphorothioate backbone. The ASOs had five nucleotides at the 5' and 3' ends modified with 2'-O-methoxyethyl (MOE) for increased stability. ASOs and siRNA sequences are provided in Supplementary Table 1. siRNA and ASOs were introduced into mouse T_H17 cells by Amara nucleofection as previously described⁸.

Wild-type and helicase-dead mutant DDX5 were described previously⁵⁴. DDX5 and *Rmrp* were subcloned into the mouse stem-cell virus (MSCV) Thy1.1 vectors for retroviral overexpression and rescue assays in T cells. Retrovirus production was carried out in Plat-E cells (Cell Biolabs, Inc., not tested for mycoplasma) as previously described⁵⁵. Spin transduction was performed 24 h after *in vitro* T-cell activation by centrifugation in a Sorvall Legend RT at 700 g for 90 min at 32 °C. Aqua⁺Thy1.1⁺ live and transduced cells were analysed by flow cytometry after 5 days of culture in T_H17-polarizing conditions.

RORγt transcriptional activity in polarized T_H17 cells. A ROR luciferase reporter was constructed with four RORE sites replacing the Gal4 (UAS) sites from the pGL4.31 vector (luc2P/GAL4 UAS/Hygro) from Promega (C935A) as described in ref. 56. Naive CD4⁺ T cells were cultured in T_H17-polarizing conditions for 72 h. Nucleofection (Amara Nucleofector 4D, Lonza) was then used to introduce 1 µg RORE-firefly luciferase reporter construct and 1 µg control *Renilla* luciferase construct according to the manufacturer's instructions. Luciferase activity was measured using the dual luciferase reporter kit (Promega) at 24 h after transfection. Relative luciferase units (RLU) were calculated as a function of firefly luciferase reads over those of *Renilla* luciferase. DMSO or 2 µM RORγ inhibitor (ML209) were used in luciferase experiments as described in ref. 57.

Co-immunoprecipitation and mass spectrometry. Cultured T_H17 cells (100 × 10⁶) were lysed in 25 mM Tris (pH 8.0), 100 mM NaCl, 0.5% NP-40, 10 mM MgCl₂, 10% glycerol, 1 × protease inhibitor and PhosphoSTOP (Roche) on ice for 30 min, followed by homogenization with a 25-gauge needle. The RORγt-specific antibody used for pull-down assays was previously described⁸. Co-immunoprecipitated complexes were collected with protein G dynabeads (Dyna, Invitrogen). Detailed antibody information is provided in Supplementary Table 1. Mass spectrometry and the Mascot database search to identify protein complex composition were both performed by the Central Proteomics Facility at the Dunn School of Pathology, Oxford, UK.

Ribosome TRAP-seq, RIP-seq and RNA-seq. Twenty million cells cultured in T_H17-polarizing conditions for 48 h were lysed in 10 mM HEPES (pH 7.4), 150 mM KCl, 5 mM MgCl₂, 0.5 mM dithiothreitol (DTT), 100 µg ml^{−1} cycloheximide, 1% NP-40, 30 mM DHPG, 1 × protease inhibitor and PhosphoSTOP (Roche). Ribosome-TRAP immunoprecipitation was first performed using 2 µg of anti-GFP antibody (Invitrogen) and collected in 20 µl of protein G magnetic dynabeads. The supernatant was removed for subsequent RIP pull-down using anti-DDX5 (Abcam) or anti-RORγt antibodies and collected with protein G dynabeads. TRAP-seq samples were washed with high-salt wash buffer (10 mM HEPES (pH 7.4), 350 mM KCl, 5 mM MgCl₂, 1% NP-40, 0.5 mM DTT and 100 µg ml^{−1} cycloheximide). RIP-seq samples were washed three times with 25 mM Tris (pH 8.0), 100 mM NaCl, 0.5% NP-40, 10 mM MgCl₂, 10% glycerol, 1 × protease inhibitor and PhosphoSTOP (Roche). Enrichment of target proteins was confirmed by immunoblot analysis. Complementary DNAs (cDNAs) were synthesized from TRIzol (Invitrogen)-isolated RNA, using Superscript III kits (Invitrogen). RNA-seq libraries were prepared and sequenced at Genome Services Laboratory, HudsonAlpha. Sequencing reads were mapped by Tophat and transcripts called by Cufflinks. Pull-down enrichment was calculated for each transcript as a ratio of FPKM recovered from TRAP-seq and RIP-seq samples compared to those from 5% input.

For RNA-seq analysis, volcano scores for wild-type, DDX5-T and RORγt-knockout T_H17 cells were calculated for each transcript as a function of its *P* value and fold change between mutant and wild-type controls. BAM files were converted to .tbf format for viewing with the IGV Browser Tool. Ingenuity Pathway Analysis (IPA, Qiagen) was used to identify enriched Gene Ontology terms in the DDX5-RORγt coregulated gene set.

ChIRP-seq and ChIRP-qPCR. The ChIRP-seq assay was performed largely as described previously⁵⁸. Mouse T_H17 cells were cultured as above and *in vivo* RNA-chromatin interactions were fixed with 1% glutaraldehyde for 10 min at 25 °C. Antisense DNA probes (designated 'odd' or 'even') against *Rmrp* were designed

by Biosearch Probe Designer (1, 5'-TAGGAAACAGGCCTTCAGAG-3'; 2, 5'-AACATGTCCTCGTAGTAG-3'; 3, 5'-CCCCTAGGCGAAAGGATAAG-3'; 4, 5'-AACAGTGACTTGCGGGGGGAA-3'; 5, 5'-CTATGTGAGCTGACGG ATGA-3'). Probes modified with BiotinTEG at the 3' end were synthesized by Integrated DNA Technologies (IDT). Isolated RNA was used in RT-qPCR analysis (Stratagene) to quantify enrichment of *Rmrp* and depletion of other cellular RNAs. Isolated DNA was used for qPCR analysis or to make deep sequencing libraries with the NEBNext DNA library prep master mix set for Illumina (NEB). Library DNA was quantified on the high sensitivity bioanalyzer (Agilent) and sequenced from a single end for 75 cycles on an Illumina NetSeq 500.

Sequencing reads were first trimmed of adaptors (FASTX Toolkit) and then mapped using Bowtie to a custom bowtie index containing single-copy loci of repetitive RNA elements (ribosomal RNAs, small nuclear RNAs, and noncoding Y RNAs⁵⁹). Reads that did not map to the custom index were then mapped to mm9. Mapped reads were separately shifted towards the 3' end using MACS and normalized to a total of 10 million reads. Even and odd replicates were merged as described previously⁵⁸ by taking the lower of the two read density values at each nucleotide across the entire genome. These processing steps take raw FASTQ files and yield processed files that contain genome-wide *Rmrp*-occupied chromatin association maps, where each nucleotide in the genome has a value that represents the relative binding level of the *Rmrp* RNA. MACS parameters were as follows: band width = 300; model fold = 10, 30; P-value cutoff = 1×10^{-5} . The full pipeline is available at <https://github.com/bdo311/chirpseq-analysis>.

ChIRP-qPCR was performed on DNA purified after treatment with RNase (60 min, 37 °C) and proteinase K (45 min, 65 °C). The primers used for qPCR are listed in Supplementary Table 1. For enrichment analysis, we tested for the enrichment of *Rmrp* ChIRP peaks among ChIP peak sets for key T_H17 transcription factors, CTCF, RNA Pol II and several histone marks⁸. Assay for transposase-accessible chromatin sequencing (ATAC)-seq was performed, according to published protocol⁶⁰, on cultured T_H17-polarized cells *in vitro* for 48 h (unpublished data). Because of differences in ChIP antibody affinities and the bias in the selection of ChIP and ChIRP factors, we used peaks generated from ATAC-seq data as a background setting for the enrichment analysis. In our analysis, we considered all ChIRP and ChIP peaks that fell within ± 500 base pairs of ATAC-seq peaks, and then calculated the overlap among the ChIRP and ChIP sets, using the hypergeometric distribution to estimate significance.

In vitro binding assay. For *in vitro* binding assays, pcDNA3.1-*Rmrp* vectors were used for T7 polymerase-driven *in vitro* transcription (IVT) reactions (Promega). Haemagglutinin (HA)-DDX5 and Flag-ROR γ t were *in vitro* transcribed and translated using an *in vitro* transcription and translation (TNT) system according to the manufacturer's protocol (Promega). Alternatively, pGEX4.1-DDX5 (wild-type and helicase-dead mutant) constructs were transformed into BL21 to synthesize recombinant full-length GST-hDDX5 proteins. Full-length His-tagged human ROR γ t was purified in three steps through Ni-resin, S column and gel-filtration (AKTA). Then, 0.5 μ g of each recombinant protein was incubated in the presence or absence of 200 μ M ATP, 300 ng *in vitro* transcribed *Rmrp* in co-immunoprecipitation buffer containing 25 mM Tris (pH 8.0), 100 mM NaCl, 0.5% NP-40, 10 mM MgCl₂, 10% glycerol, 1 \times protease inhibitor, RNaseInhibitor (Invitrogen) and PhosphoSTOP (Roche). GST-DDX5 was enriched on glutathione beads (GE); HA-DDX5, Flag-ROR γ t and His-ROR γ t were enriched using anti-HA (Covance), anti-Flag (Sigma) and anti-His antibodies (Santa Cruz Bio) coupled to anti-mouse immunoglobulin dynabeads (Dyna, Invitrogen).

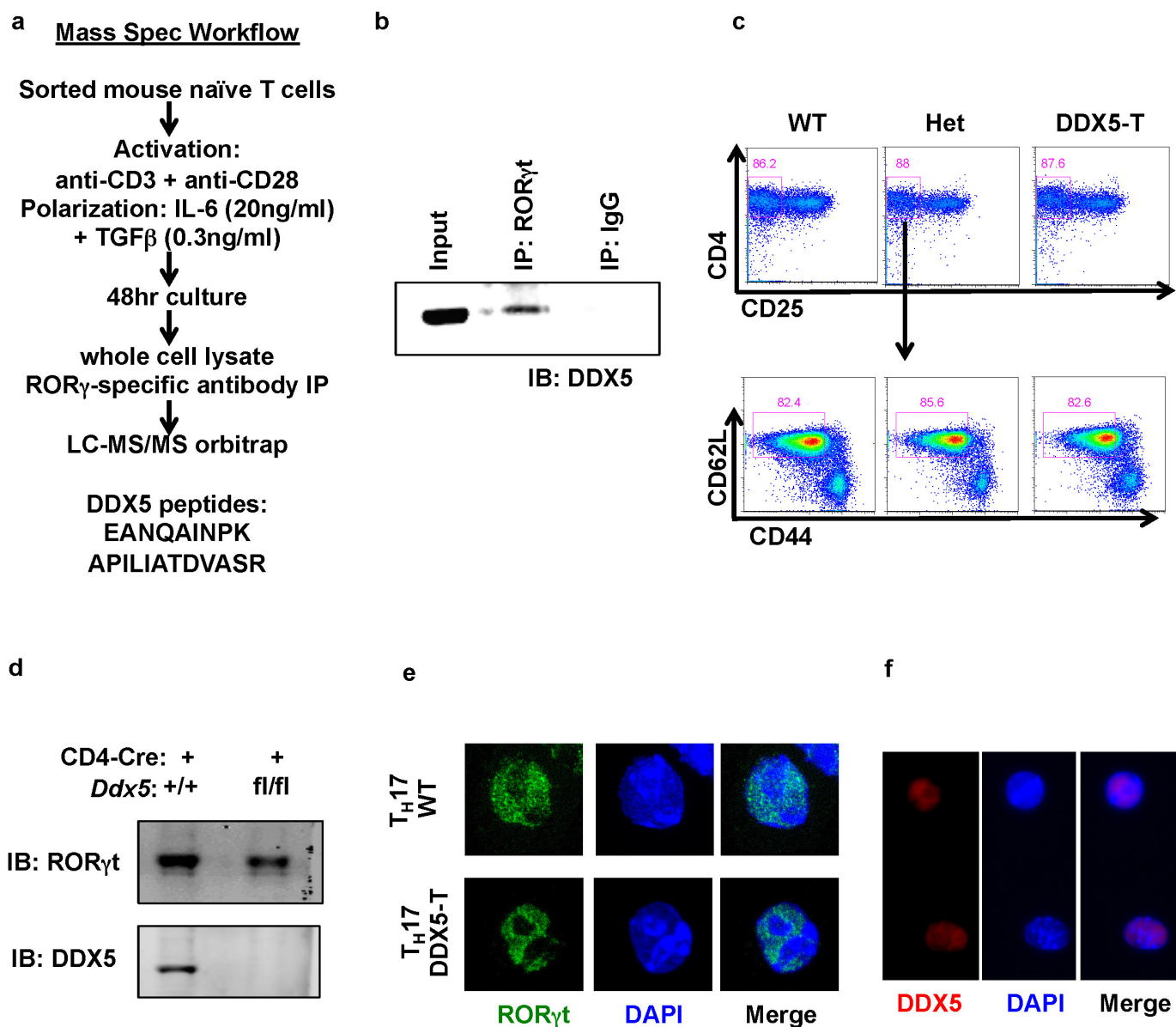
Microscopy. T_H17 cells were cultured on glass coverslips for 48 h and fixed in 4% paraformaldehyde in PBS for 5 min at room temperature. Fixed cells were permeabilized with 0.1% bovine serum albumin (BSA), 0.1% Triton and 10% normal serum in PBS for 1 h. Cells were then incubated with primary antibodies (DDX5

(Abcam) or ROR γ t (eBiosciences)) in 0.1% BSA and 0.2% Triton PBS overnight at 4 °C. Secondary antibodies (anti-goat Alexa 488 or anti-rat Alexa 647 (Molecular Probe)) were incubated at 4 °C for 1 h. Stained cells were washed three times with 0.5% Tween and 0.1% BSA in PBS. DAPI was used to stain DNA inside the nucleus. Immunofluorescence images were captured on a Zeiss 510 microscope at 40 \times .

ChIP and RT-qPCR analysis. T_H17-polarized cells were crosslinked with 1% paraformaldehyde (EMS) and incubated with rotation at room temperature. Crosslinking was stopped after 10 min with glycine to a final concentration of 0.125 M and incubated for a further 5 min with rotation. Cells were washed with 3 \times ice-cold PBS and pellets were either flash-frozen in liquid N₂ or immediately resuspended in Farnham lysis buffer (5 mM PIPES, 85 mM KCl, 0.5% NP-40). Hypotonic lysis continued for 10 min on ice before cells were spun down and resuspended in RIPA buffer (1 \times PBS, 1% NP-40, 0.5% SDS, 0.5% Na-deoxycholate), transferred into TPX microtubes and lysed on ice for 30 min. Nuclear lysates were sonicated for 40 cycles of 30 s 'ON' and 30 s 'OFF' in 10-cycle increments using a Biorupter (Diadene) set on high. After pelleting debris, chromatin was pre-cleared with protein G dynabeads (dynabeads, TFS) for 2 h with rotation at 4 °C. For immunoprecipitation, pre-cleared chromatin was incubated with anti-ROR γ t antibodies (1 μ g per 2 $\times 10^6$ cells) overnight with rotation at 4 °C and protein G was added for the final 2 h of incubation. Beads were washed and bound chromatin was eluted. ChIP-qPCR was performed on DNA purified after treatment with RNase (30 min, 37 °C) and proteinase K (2 h, 55 °C) followed by reversal of crosslinks (8–12 h, 65 °C). The primers used for qPCR have been described previously⁸.

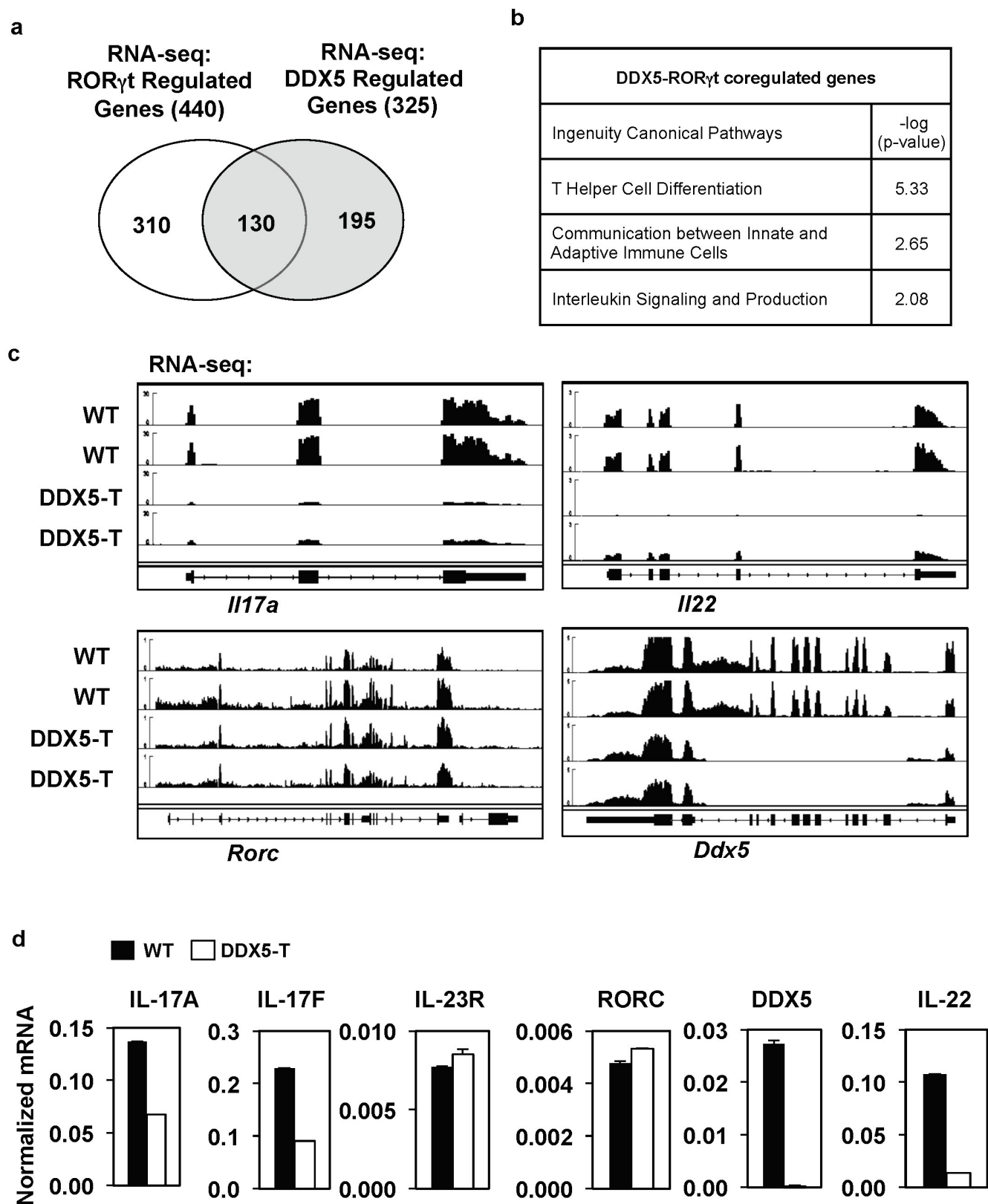
For analysis of mRNA transcripts, gene specific values were normalized to the *Gapdh* housekeeping gene for each sample. All primer sequences are listed in Supplementary Table 1.

47. Stanley, S. *et al.* Profiling of glucose-sensing neurons reveals that GHRH neurons are activated by hypoglycemia. *Cell Metab.* **18**, 596–607 (2013).
48. Nicol, S. M. *et al.* The RNA helicase p68 (DDX5) is selectively required for the induction of p53-dependent p21 expression and cell-cycle arrest after DNA damage. *Oncogene* **32**, 3461–3469 (2013).
49. Crosswell, A., Amir, E., Tegatz, P., Barman, M. & Salzman, N. H. Prolonged impact of antibiotics on intestinal microbial ecology and susceptibility to enteric *Salmonella* infection. *Infect. Immun.* **77**, 2741–2753 (2009).
50. Ostanin, D. V. *et al.* T cell transfer model of chronic colitis: concepts, considerations, and tricks of the trade. *Am. J. Physiol. Gastrointest. Liver Physiol.* **296**, G135–G146 (2009).
51. Kim, S. V. *et al.* GPR15-mediated homing controls immune homeostasis in the large intestine mucosa. *Science* **340**, 1456–1459 (2013).
52. Manel, N., Unutmaz, D. & Littman, D. R. The differentiation of human T_H-17 cells requires transforming growth factor- β and induction of the nuclear receptor ROR γ mat. *Nature Immunol.* **9**, 641–649 (2008).
53. Meng, L. *et al.* Towards a therapy for Angelman syndrome by targeting a long non-coding RNA. *Nature* **518**, 409–412 (2015).
54. Bates, G. J. *et al.* The DEAD box protein p68: a novel transcriptional coactivator of the p53 tumour suppressor. *EMBO J.* **24**, 543–553 (2005).
55. Morita, S., Kojima, T. & Kitamura, T. Plat-E: an efficient and stable system for transient packaging of retroviruses. *Gene Ther.* **7**, 1063–1066 (2000).
56. Santori, F. R. *et al.* Identification of natural ROR γ ligands that regulate the development of lymphoid cells. *Cell Metab.* **21**, 286–297 (2015).
57. Huh, J. R. *et al.* Identification of potent and selective diphenylpropanamide ROR γ inhibitors. *ACS Med. Chem. Lett.* **4**, 79–84, (2013).
58. Chu, C., Qu, K., Zhong, F. L., Artandi, S. E. & Chang, H. Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell* **44**, 667–678 (2011).
59. Flynn, R. A. *et al.* Dissecting noncoding and pathogen RNA-protein interactomes. *RNA* **21**, 135–143 (2015).
60. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1–21.29.9. (2015).



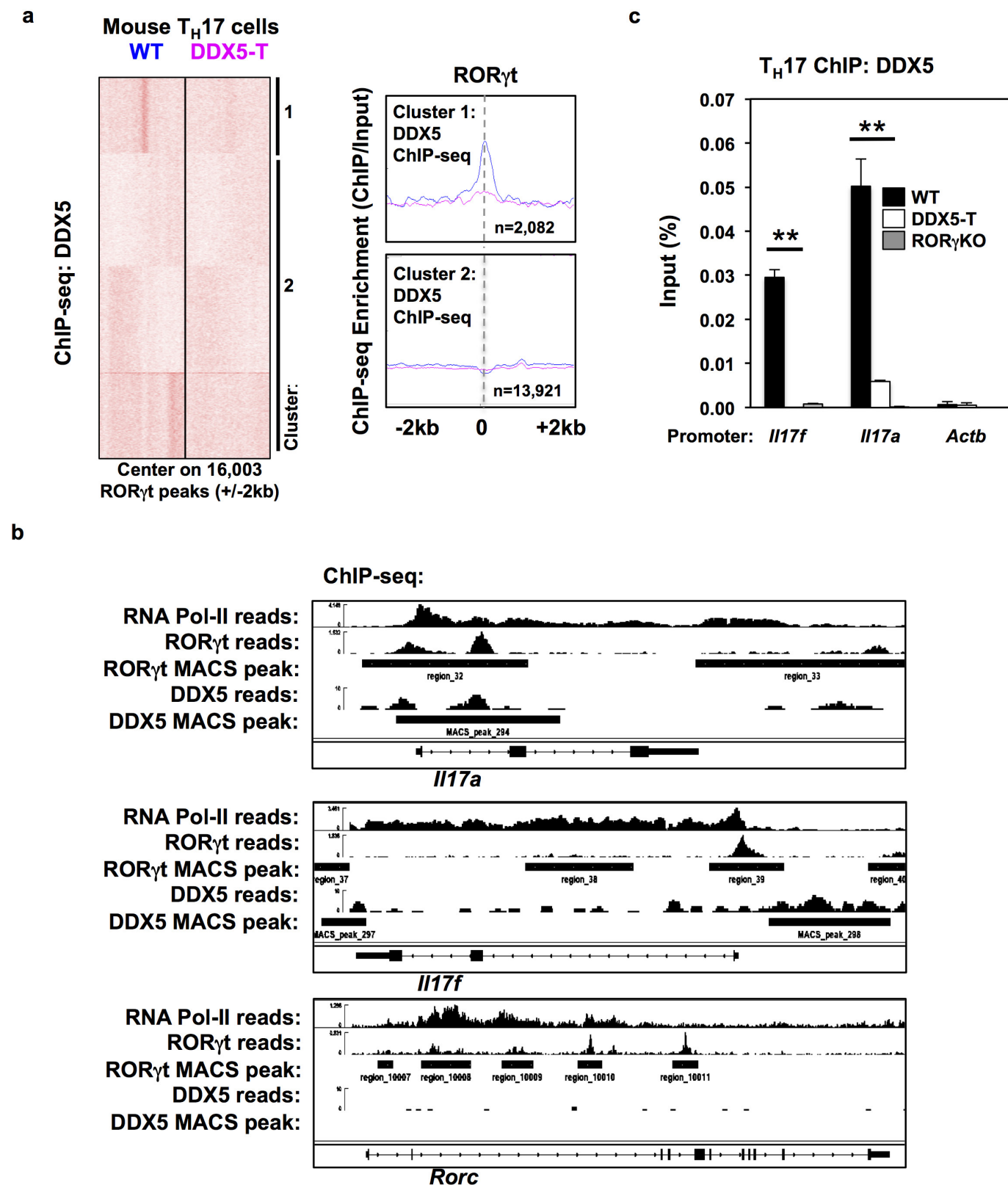
Extended Data Figure 1 | Identification of DDX5 as a RORγt-interacting partner. **a**, Mass spectrometry experimental workflow. Sorted naïve CD4⁺ T cells from wild-type mice were cultured *in vitro* in T_H17-polarizing conditions for 48 h. Immunoprecipitation of endogenous RORγt was performed using RORγt-specific antibodies on whole-cell lysates. RORγt enrichment in pull-down was confirmed by immunoblot. Immunoprecipitated proteins were digested and analysed by mass spectrometry. The listed DDX5 peptides were identified in the T_H17 RORγt immunoprecipitate. **b**, Co-immunoprecipitation of DDX5

with anti-RORγt in lysates of *in vitro* polarized T_H17 cells. **c**, Cell surface phenotype of splenic and lymph node DAPI⁺ CD4⁺ CD8α⁻ CD19⁻ T cells from wild-type and DDX5-T mice, examined by flow cytometry. **d**, Immunoblot of RORγt protein expression whole-cell lysate of cultured T_H17 cells from wild-type or DDX5-T animals. For uncropped gels (**b**, **d**), see Supplementary Fig. 1. **e**, Immunofluorescence staining of RORγt in cultured T_H17 cells from wild-type or DDX5-T mice. **f**, Immunofluorescence staining revealed nuclear localization of DDX5 in T_H17 cells.



Extended Data Figure 2 | DDX5 coregulates a subset of ROR γ t transcriptional targets in polarized T_H17 cells. **a**, Venn diagram of distinct and overlapping genes regulated by ROR γ t and/or DDX5, as determined from RNA-seq studies. **b**, Ingenuity Pathway Analysis (Qiagen) of DDX5- and ROR γ t-coregulated genes. **c**, IGV browser view

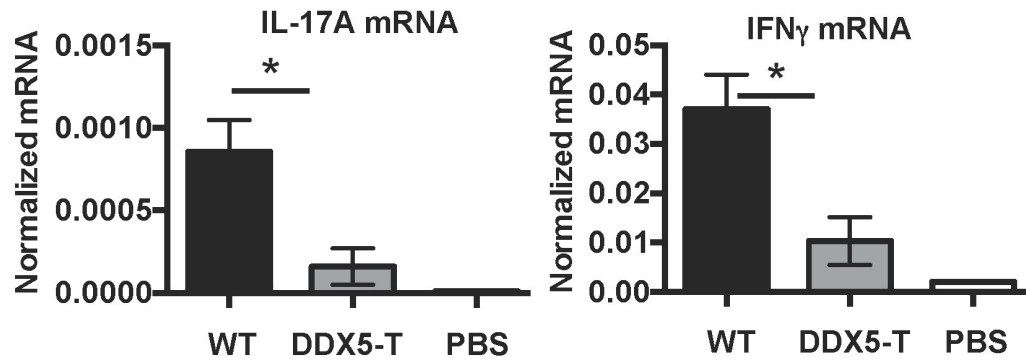
showing biological replicate RNA-seq coverage tracks of control and DDX5-T from *in vitro* polarized T_H17 cell samples at the *Il17a*, *Il22*, *Ddx5* and *Rorc* loci. **d**, Independent RT-qPCR validation of RNA-seq results confirming effects of DDX5 deletion on ROR γ t target gene expression. Graphs show mean \pm s.d.



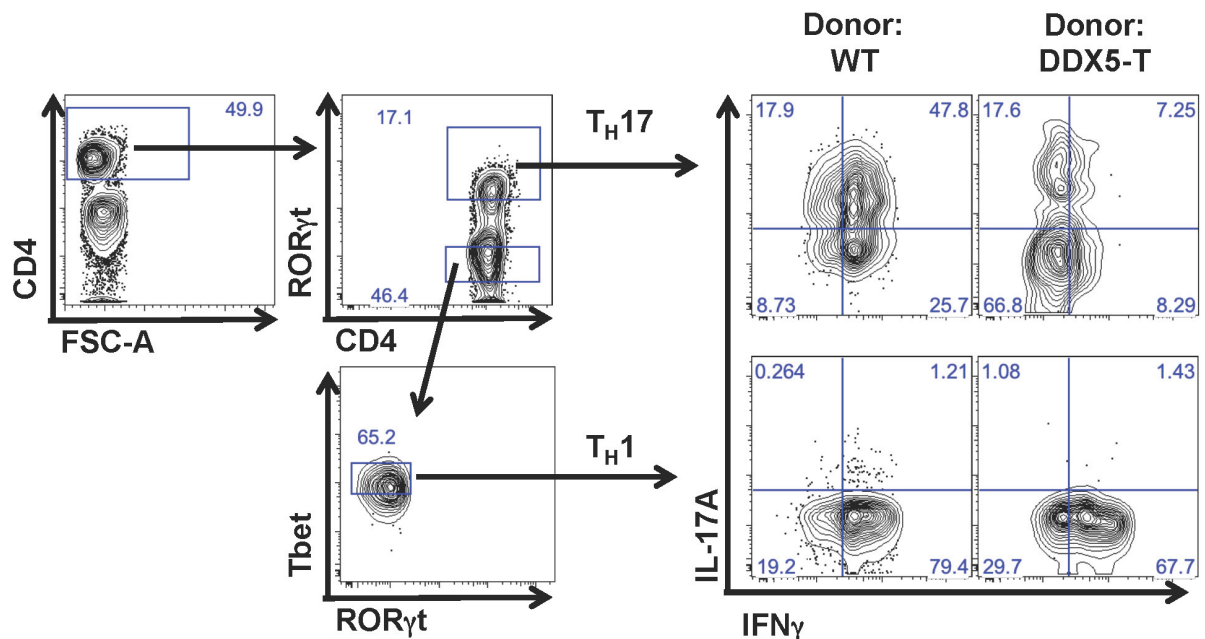
Extended Data Figure 3 | DDX5 chromatin localization in T_H17 cells. a, ChIP-seq-generated heatmap of DDX5 occupancy in regions centred on 16,003 ROR γ t-occupied sites (\pm 2 kilobases (kb)). K-means linear normalization was used for clustering analysis by SeqMiner. Metagene analysis on cluster 1 depicts ROR γ t-occupied regions with DDX5 enrichment in wild-type but not DDX5-T cells; cluster 2 represents ROR γ t-occupied regions without DDX5 enrichment. b, IGV browser view of *Il17a*, *Il17f* and *Rorc* loci with ChIP-seq enrichment for RNA Pol II,

ROR γ t and DDX5. c, Independent ChIP-qPCR of DDX5 in polarized T_H17 cells. DDX5 occupancy at the *Il17a* and *Il17f* loci (as identified by ROR γ t ChIP-seq MACS peak called 32 and 39, respectively, from b) in control, DDX5-T or ROR γ -deficient (ROR γ KO) cells. Results are representative of two independent experiments. Each experiment was performed with two technical replicates. Graph shows mean \pm s.d. ** $P < 0.01$ (unpaired, *t*-test).

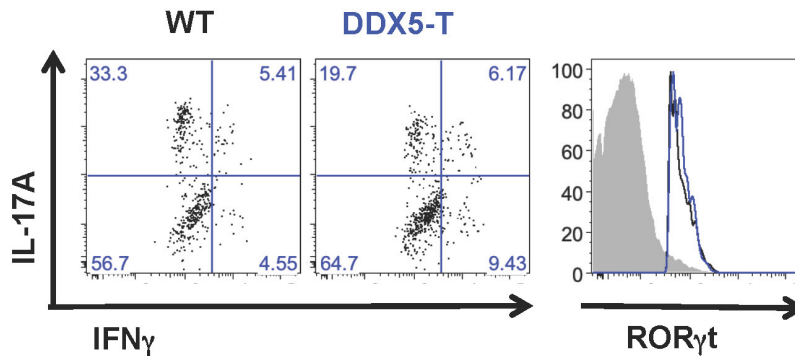
a



b

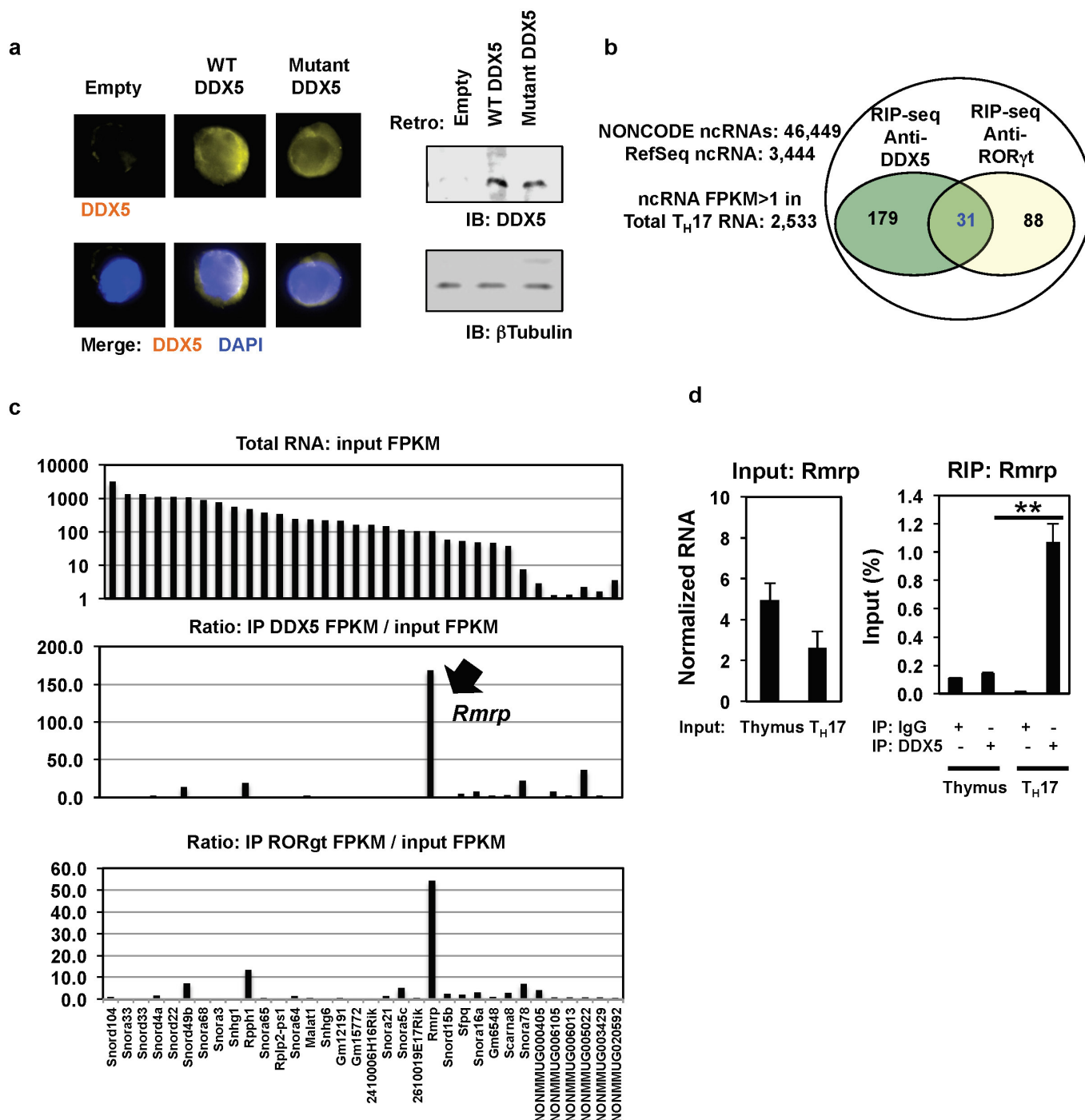


c



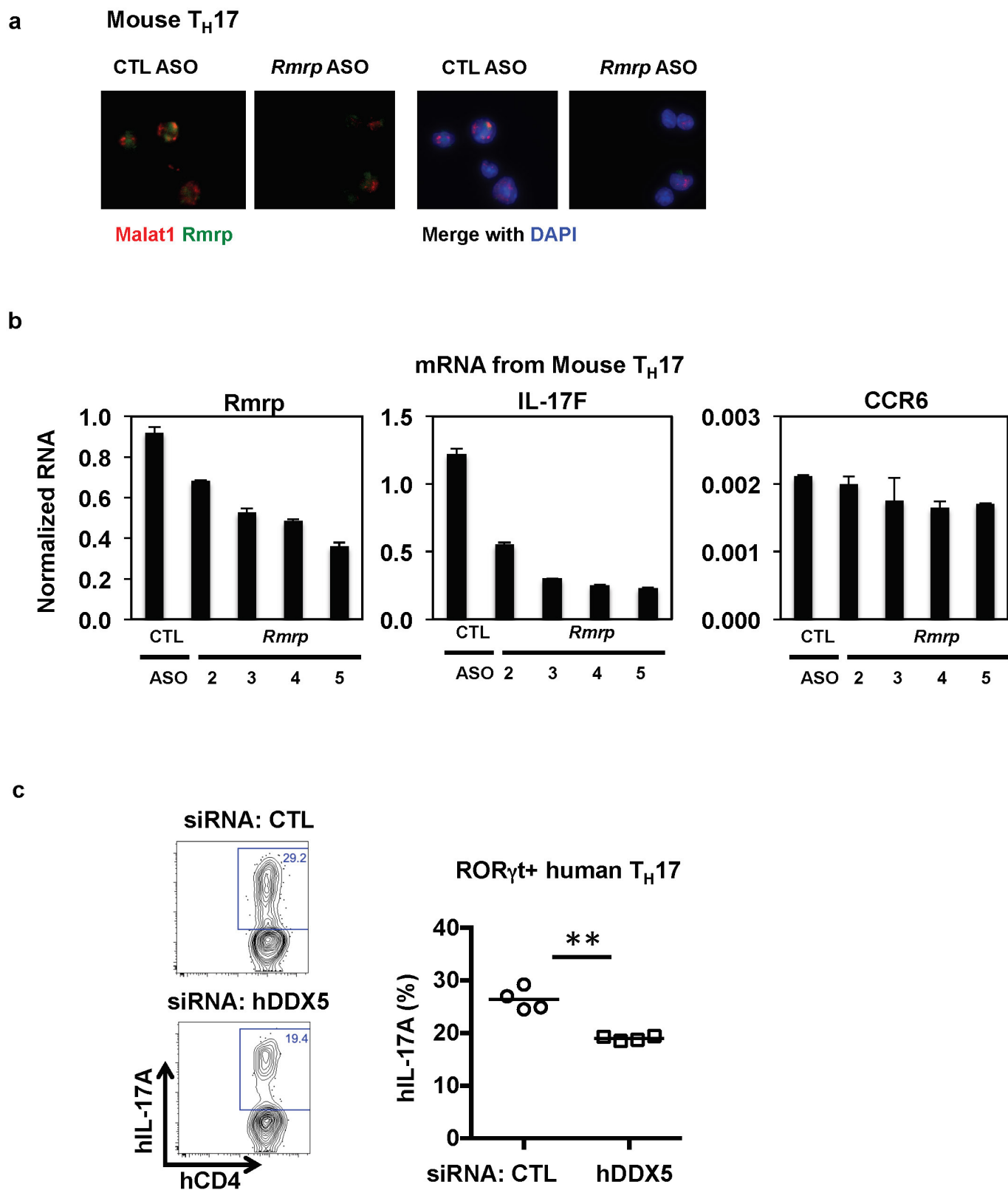
Extended Data Figure 4 | Influence of DDX5 on T-cell phenotypes in autoimmune disease models. **a**, At 8 weeks after T-cell transfer, large intestine lamina propria mononuclear cells were evaluated for amounts of *Il17a* and *Ifng* mRNA by RT-qPCR. Results are representative of two independent experiments. Each experiment was performed using large intestines from three mice in each condition. RT-qPCR was performed

with two technical replicates. Graph shows mean \pm s.d. * $P < 0.03$ (unpaired, t -test). **b**, Gating strategy for analysis of T_H17 and T_H1 cells from large intestine of Rag2-deficient recipients of wild-type or DDX5-T naive T cells analysed at 8 weeks after T-cell transfer. **c**, Representative IL-17A and IFN γ intracellular staining of Aqua⁻CD4⁺ROR γ t⁺ T_H17 cells in spinal cord of MOG-immunized animals on day 21.



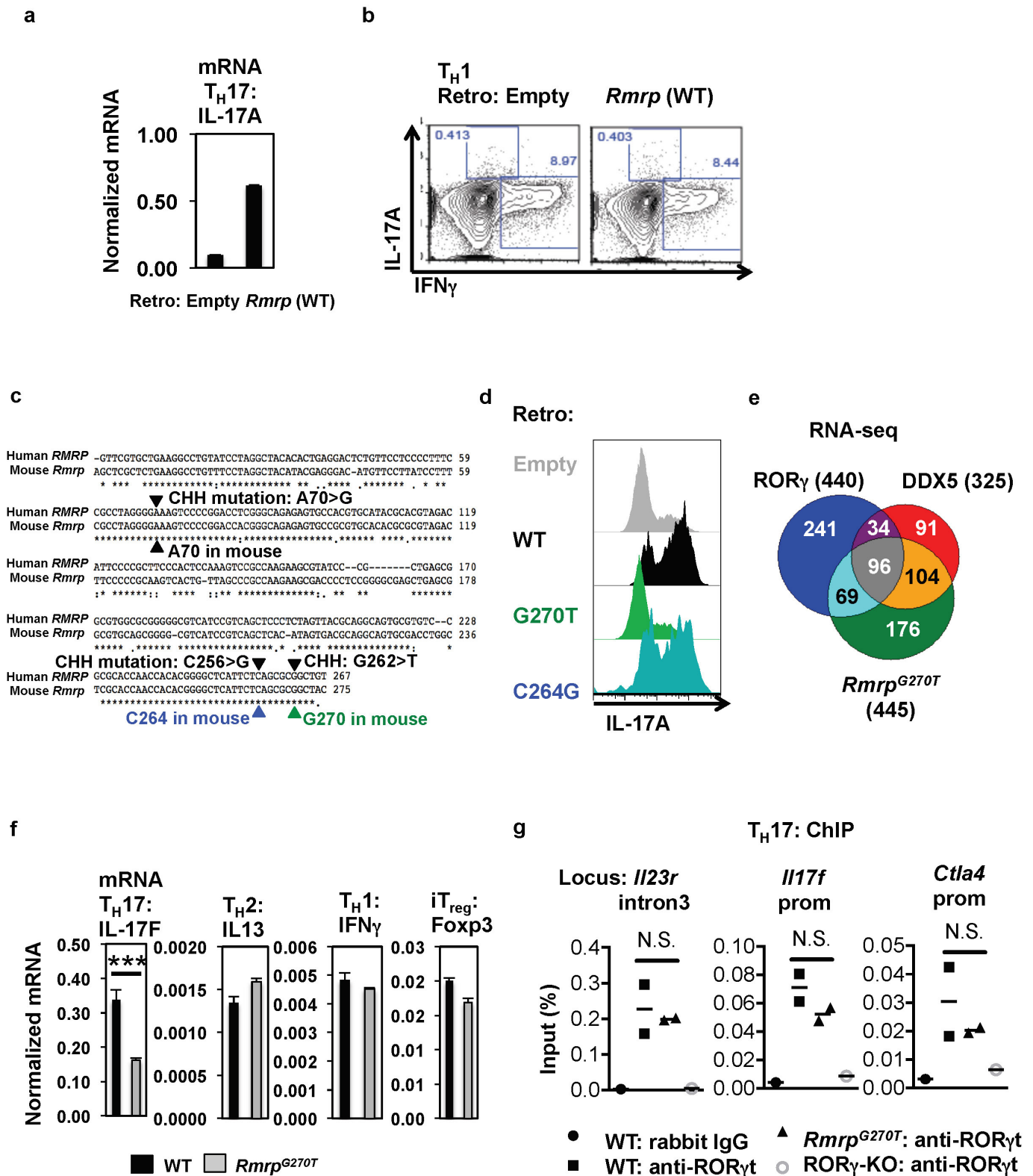
Extended Data Figure 5 | Noncoding RNAs enriched in DDX5 and ROR γ t RIP-seq studies. **a**, DDX5- T_H17 cells were transduced with wild-type or helicase-mutant DDX5 and evaluated for DDX5 expression by immunofluorescence (left) and immunoblot (right) with anti-DDX5 antibody. For uncropped gels, see Supplementary Fig. 1. **b**, Venn diagram of noncoding RNAs detected by RIP-seq of ribosome-depleted T_H17 cell lysates with anti-DDX5 and anti-ROR γ t antibodies. **c**, Abundance of top

noncoding RNAs enriched in DDX5 and ROR γ t immunoprecipitates from polarized T_H17 cell lysates depleted of ribosomes. Top, abundance of the noncoding RNAs in total lysate. **d**, RIP-qPCR experiments to compare *Rmrp* association with DDX5 in cultured T_H17 and total thymocytes *ex vivo*. Results are representative of three independent experiments. Each experiment was performed with two technical replicates. Graph shows mean \pm s.d. ** $P < 0.001$ (unpaired, *t*-test).



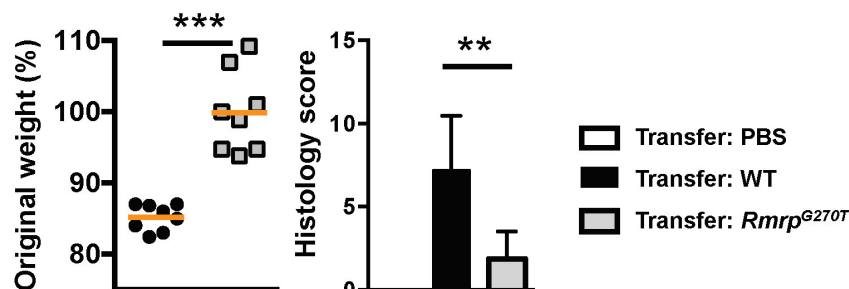
Extended Data Figure 6 | *Rmrp* and DDX5 knockdown in mouse and human T_H17 cells. **a**, RNA FISH analysis, using probes specific for *Rmrp* (green) and *Malat1* (red) lncRNAs, in T_H17 cells at 72 h after nucleofection with control (CTL) or *Rmrp* ASOs. **b**, Effect of *Rmrp* ASOs targeting different regions of *Rmrp* transcript on levels of *Rmrp*, *Il17f* and *Ccr6*

RNAs in polarized T_H17 cells. **c**, Knockdown of DDX5 reduced IL-17A production in *in vitro* polarized human ROR γ t⁺ T_H17 cells. ** $P < 0.01$ (paired, *t*-test). Representative result shown in left panel. Each dot represents a different healthy donor ($n = 4$). Graphs show mean \pm s.d.

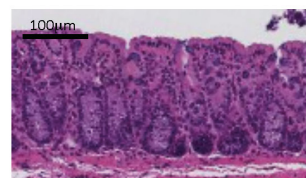
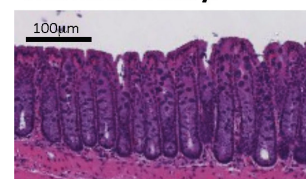
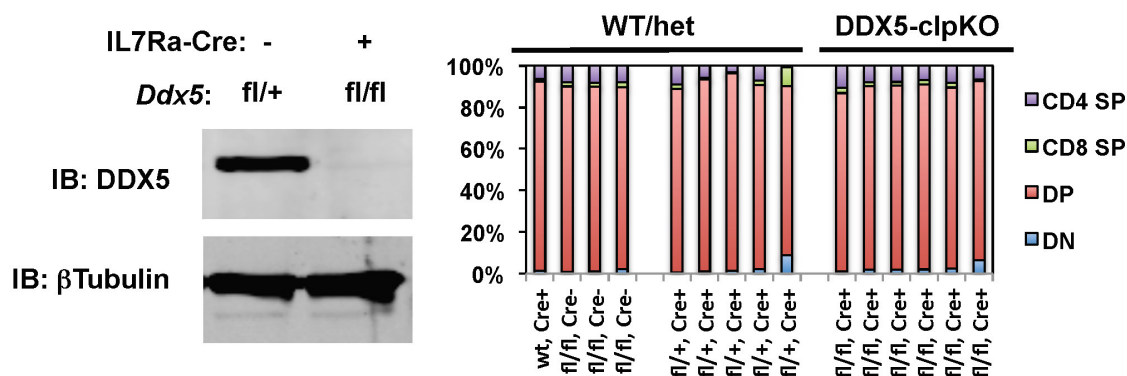
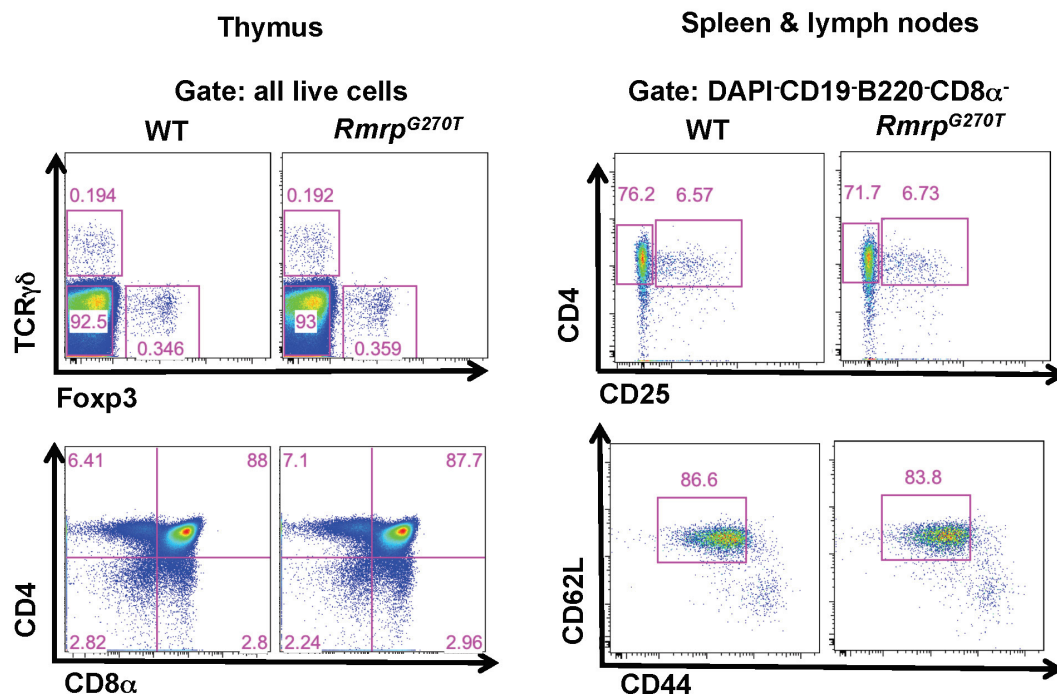


Extended Data Figure 7 | Effects of wild-type and mutant *Rmrp* in T-cell differentiation. **a**, *Il17a* mRNA in cell lysates of *in vitro* polarized mouse T_H17 cells at 96 h after transduction of control vector or wild-type *Rmrp*. Results are representative of two independent experiments. **b**, IFN_γ production in polarized mouse T_H1 cells at 96 h after transduction of control or *Rmrp*-encoding vector. Representative of two independent experiments. Each experiment was performed with two technical replicates. **c**, Comparison of human and mouse *Rmrp* sequences. Several mutations identified in CHH patients are highlighted. **d**, IL-17A production in polarized mouse T_H17 cells at 96 h after transduction of wild-type or mutant *Rmrp* vectors. Representative of two independent experiments. **e**, Venn

diagram depicting the number of distinct and overlapping genes regulated by ROR_γt, DDX5 and *Rmrp* in *in vitro* polarized T_H17 cells. **f**, Expression of cytokine and *Foxp3* mRNAs in T cells from wild-type or *Rmrp*^{G270T/G270T} mice cultured *in vitro* in T_H17-, iT_{reg}-, T_H1- and T_H2-polarizing conditions. Results are representative of two independent experiments. Each experiment was performed with two technical replicates. ****P* < 0.001 (unpaired, *t*-test). **g**, ChIP-qPCR experiment using anti-ROR_γ/γt antibodies on chromatin of T_H17 cells from wild-type or mutant mice cultured for 48 h *in vitro*. Each dot represents a different biological sample. Wild type, *n* = 2; *Rmrp*^{G270T}, *n* = 2. Results are representative of three separate independent experiments. Graphs show mean ± s.d. (unpaired, *t*-test).

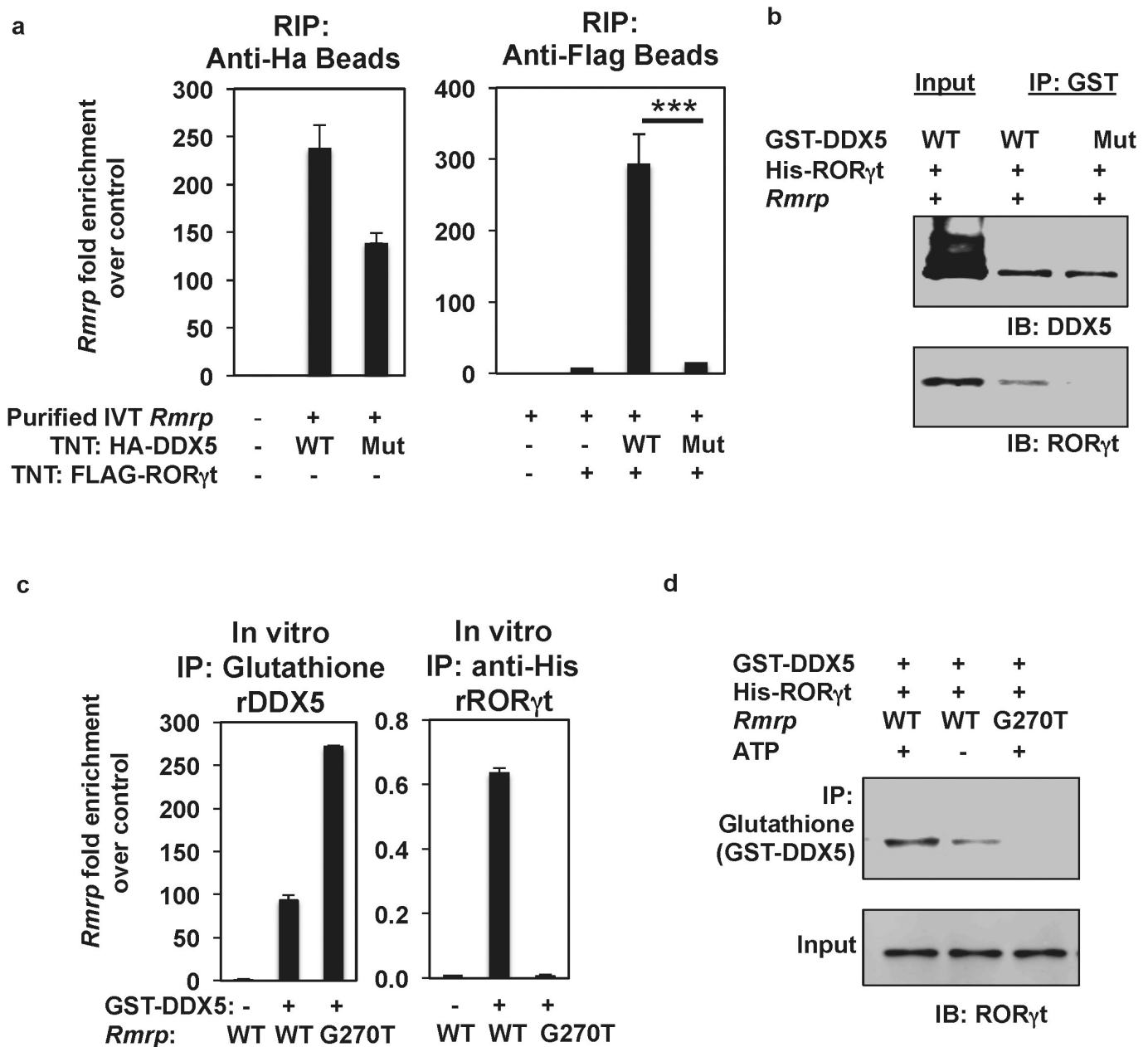
a 8 wks after transfer

Transfer: WT

Transfer: *Rmrp*^{G270T}**b****c**

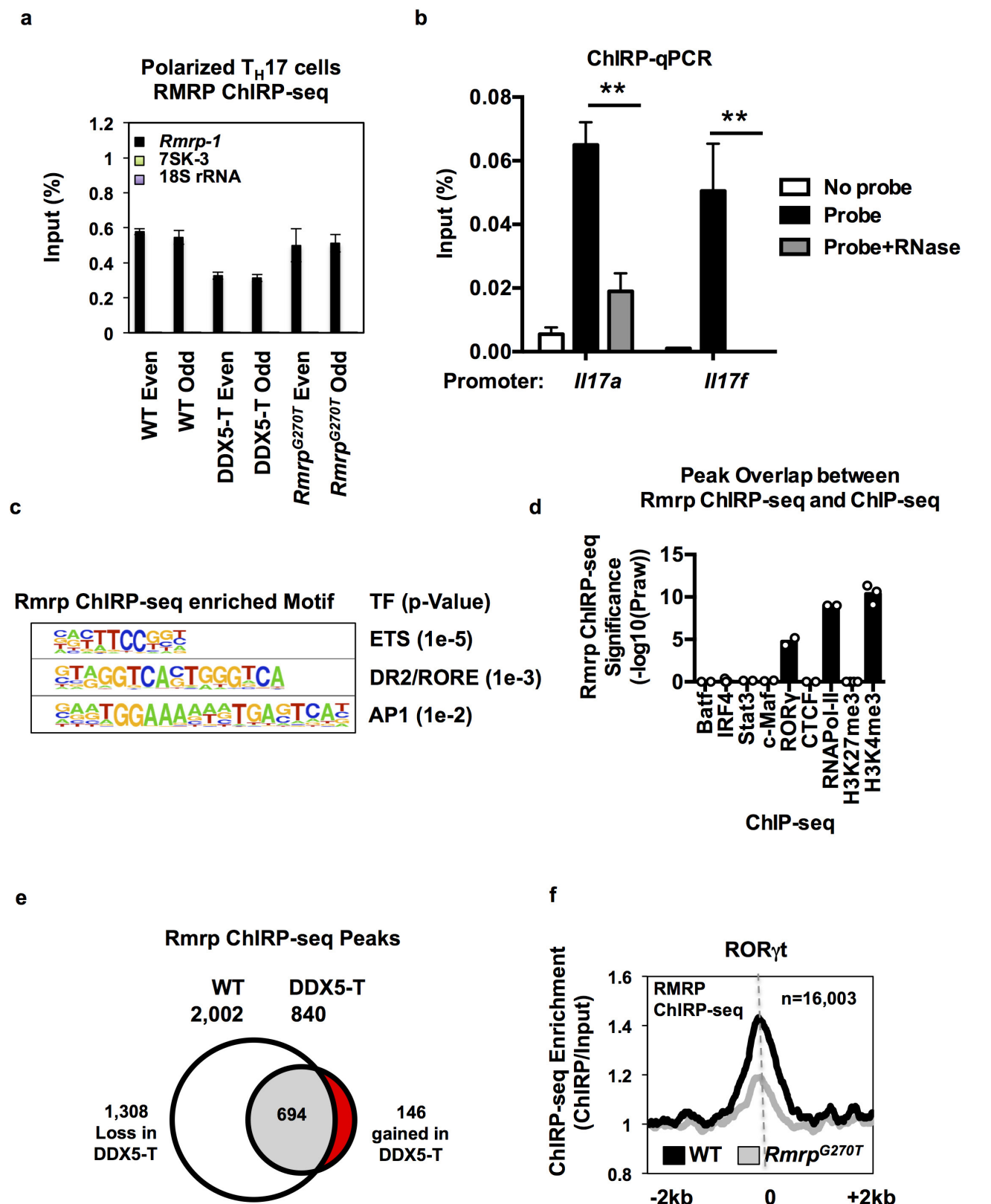
Extended Data Figure 8 | Effect of *Ddx5* and *Rmrp* mutations in inflammation and thymocyte development. **a**, Left, percentage weight change in *Rag2*^{-/-} recipients of wild-type (black circles) or *Rmrp*^{G270T/G270T} (grey squares) naive CD4⁺ T cells in the transfer model of colitis. Animal weight was measured on day 56 (wild type, *n* = 8; *Rmrp*^{G270T/G270T}, *n* = 8, combined from three independent experiments). Graphs show mean ± s.d. ****P* < 0.001 (unpaired, *t*-test). Middle, histology score (scale of 0–24) (wild type, *n* = 8; *Rmrp*^{G270T/G270T}, *n* = 5), combined from two independent experiments. ***P* < 0.01 (unpaired, *t*-test). Right, representative H&E staining of large intestine from *Rag2*^{-/-} mice on day 56 after naive T-cell

transfer. **b**, Mice with deletion of *Ddx5* in early common lymphoid progenitors (DDX5-clpKO) have normal thymic development. Left, immunoblot of thymocyte lysates with anti-DDX5 antibody confirmed depletion of DDX5. Right, percentage of CD4 single-positive (SP), CD8α SP, double-positive (DP) and double-negative (DN) cells among total thymocytes. Each bar represents the result from one mouse (WT/het, *n* = 9; DDX5-clpKO, *n* = 6). For uncropped gels, see Supplementary Fig. 1. **c**, Thymocyte and peripheral T-cell surface phenotypes of wild-type and *Rmrp*^{G270T/G270T} knock-in mice at steady state. Peripheral T-cell gate, DAPI⁻CD19⁻CD8α⁻CD4⁺.



Extended Data Figure 9 | Association of *Rmrp* lncRNA with DDX5 and ROR γ t *in vitro*. **a**, *In vitro* translated (TNT) HA-tagged wild-type or helicase-dead DDX5 and Flag-tagged ROR γ t were incubated with *in vitro* transcribed *Rmrp*. After capture on anti-HA or anti-Flag beads, the amount of lncRNA was determined by RT-qPCR. Data are representative of two independent experiments, and each experiment was performed with two technical replicates. **b**, Helicase requirement for *in vitro* interaction of DDX5 and ROR γ t. Recombinant GST-DDX5 (wild type or helicase-dead mutant) and His-ROR γ t full-length protein were synthesized in *Escherichia coli*, purified, and assayed for binding with or without *in vitro* transcribed *Rmrp* RNA in the presence exogenous

ATP. **c**, Association of *in vitro* transcribed wild-type and mutant *Rmrp* with recombinant GST-DDX5 captured on glutathione beads (left) or with recombinant GST-DDX5 and His-ROR γ t captured with anti-His antibody. Amounts of associated *Rmrp* were quantified using RT-qPCR. Data are representative of two independent experiments. Each experiment was performed with two technical replicates. **d**, Comparison of ability of *in vitro* transcribed wild-type and *Rmrp*^{G270T} lncRNA to promote interaction between recombinant ROR γ t and DDX5 *in vitro*. All graphs show mean \pm s.d. *** P < 0.001 (unpaired, *t*-test). For uncropped gels, see Supplementary Fig. 1.



Extended Data Figure 10 | *Rmrp* chromatin localization in T_H17 cells.
a, ChIRP-seq sample validation of *Rmrp* RNA pull-down over other nuclear noncoding RNAs using pools of 'even' or 'odd' capture probes. Graphs show mean \pm s.d. **b**, ChIRP-qPCR of *Rmrp* RNA pull-down from wild-type T_H17 cell lysate treated with or without RNase ($n = 2$). qPCR for each sample was performed with two technical replicates. Graph shows mean \pm s.d. **c**, HOMER motif analysis reveals top three DNA motifs within *Rmrp*-enriched peaks. **d**, Significance of peak overlaps between *Rmrp* ChIRP-seq and ChIP-seq for BATF

($n = 2$), IRF4 ($n = 7$), STAT3 ($n = 2$), c-Maf ($n = 2$), RORγt ($n = 2$), CTCF ($n = 2$), RNA Pol II ($n = 2$), H3K27me3 ($n = 4$) and H3K4me3 ($n = 3$) in T_H17 cells (hypergeometric distribution). Each dot represents a separate biological replicate of ChIP-seq experiments. **e**, Venn diagram depicting changes in peaks called from *Rmrp* (ChIRP-seq) experiments in wild-type and DDX5-T T_H17 cells. **f**, Comparison of *Rmrp* chromatin occupancy (ChIRP-seq) at known RORγt occupied loci in *in vitro* polarized T_H17 cells from wild-type and *Rmrp*^{G270T/G270T} mice.

Dense magnetized plasma associated with a fast radio burst

Kiyoshi Masui^{1,2}, Hsiu-Hsien Lin³, Jonathan Sievers^{4,5}, Christopher J. Anderson⁶, Tzu-Ching Chang⁷, Xuelei Chen^{8,9}, Apratim Ganguly¹⁰, Miranda Jarvis¹¹, Cheng-Yu Kuo^{7,12}, Yi-Chao Li⁸, Yu-Wei Liao⁷, Maura McLaughlin¹³, Ue-Li Pen^{2,14,15}, Jeffrey B. Peterson³, Alexander Roman³, Peter T. Timbie⁶, Tabitha Voytek^{3,4} & Jaswant K. Yadav¹⁶

Fast radio bursts are bright, unresolved, non-repeating, broadband, millisecond flashes, found primarily at high Galactic latitudes, with dispersion measures much larger than expected for a Galactic source^{1–7}. The inferred all-sky burst rate⁸ is comparable to the core-collapse supernova rate⁹ out to redshift 0.5. If the observed dispersion measures are assumed to be dominated by the intergalactic medium, the sources are at cosmological distances with redshifts of 0.2 to 1 (refs 10 and 11). These parameters are consistent with a wide range of source models^{12–17}. One fast burst⁶ revealed circular polarization of the radio emission, but no linear polarization was detected, and hence no Faraday rotation measure could be determined. Here we report the examination of archival data revealing Faraday rotation in the fast radio burst FRB 110523. Its radio flux and dispersion measure are consistent with values from previously reported bursts and, accounting for a Galactic contribution to the dispersion and using a model of intergalactic electron density¹⁰, we place the source at a maximum redshift of 0.5. The burst has a much higher rotation measure than expected for this line of sight through the Milky Way and the intergalactic medium, indicating magnetization in the vicinity of the source itself or within a host galaxy. The pulse was scattered by two distinct plasma screens during propagation, which requires either a dense nebula associated with the source or a location within the central region of its host galaxy. The detection in this instance of magnetization and scattering that are both local to the source favours models involving young stellar populations such as magnetars over models involving the mergers of older neutron stars, which are more likely to be located in low-density regions of the host galaxy.

We searched for fast radio bursts (FRBs) in a data archive we collected for the Green Bank Hydrogen Telescope (GBT) Intensity Mapping survey^{18–20}. The data span the frequency range 700 MHz to 900 MHz in 4,096 spectral channels. Average spectra are recorded at 1.024-ms intervals. We developed a new tree dispersion-removal algorithm and associated computer program to search for FRBs. First we removed cold plasma dispersion, which is a frequency-dependent time delay:

$$t_{\text{delay}} = 4,148.808 \text{ s} (\text{DM}/\text{pc cm}^{-3})(\text{MHz}^2/\nu^2)$$

where ν is the radio frequency, and the dispersion measure, $\text{DM} = \int n_e dl$, is the line-of-sight integral of the free electron number density n_e . We then summed all frequency channels for DM values in the range 0–2,000 pc cm^{-3} and flagged as candidates all data sets with

8 σ positive excursions of flux. These 6,496 candidates were examined by eye and compared to synthetic DM-time images of simulated FRB events. Most of these candidates have the characteristics of radio-frequency interference but one matched the expected pattern of an FRB (see Fig. 1 and Extended Data Fig. 1). This burst, hereafter referred to as FRB 110523, has a DM of 623.30(6) pc cm^{-3} ; the maximum DM expected in this direction owing to Galactic contributions²¹ is 45 pc cm^{-3} . (Here, and throughout, the measurement uncertainties in parentheses enclose the 68% confidence interval from the model fit.) Detailed parameters for the burst are given in Extended Data Table 1.

Our detection in a distinct band and with independent instrumentation compared to the 1.4-GHz detections at the Parkes and Arecibo observatories greatly strengthens the argument that FRBs are astrophysical phenomena. In addition, as described in the Methods, the close fit to astronomical expectations of FRB 110523 for dispersion spectral index, Faraday rotation spectral index, and scattering spectral index all further support an astronomical origin.

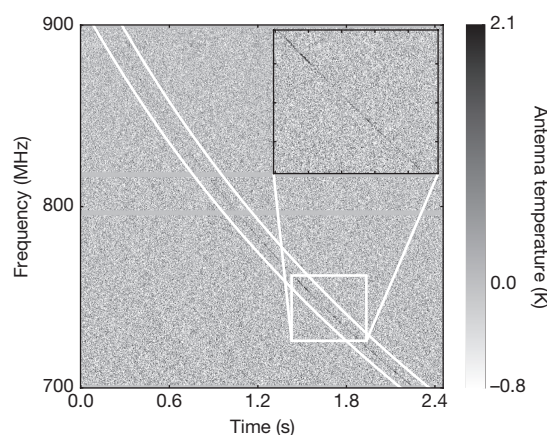


Figure 1 | Brightness temperature spectra versus time for FRB 110523.

The diagonal black curve shows the pulse of radio brightness sweeping over time. The arrival time is differentially delayed (dispersed) by plasma along the line of sight. A pair of curves in white, bracketing the FRB pulse, show that the delay function matches the one expected from cold plasma. The grey horizontal bars show where data has been omitted owing to resonances within the GBT receiver. The inset shows fluctuations in brightness caused by scintillation.

¹Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, British Columbia V6T 1Z1, Canada. ²Canadian Institute for Advanced Research, CIFAR Program in Cosmology and Gravity, Toronto, Ontario M5G 1Z8, Canada. ³McWilliams Center for Cosmology, Carnegie Mellon University, Department of Physics, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213, USA. ⁴Astrophysics and Cosmology Research Unit, School of Chemistry and Physics, University of KwaZulu-Natal, Durban 4001, South Africa. ⁵National Institute for Theoretical Physics (NITheP), KZN node, Durban 4001, South Africa. ⁶Department of Physics, University of Wisconsin, Madison, Wisconsin 53706-1390, USA. ⁷Academia Sinica Institute of Astronomy and Astrophysics, 11F Astro-Math Building, AS/NTU 1, Section 4, Roosevelt Road, Taipei 10617, Taiwan. ⁸National Astronomical Observatories, Chinese Academy of Science, 20A Datun Road, Beijing 100012, China. ⁹Center of High Energy Physics, Peking University, Beijing 100871, China. ¹⁰Astrophysics and Cosmology Research Unit, School of Mathematics, Statistics, and Computer Science, University of KwaZulu-Natal, Durban 4001, South Africa. ¹¹Department of Astronomy and Astrophysics, University of Toronto, 50 St George Street, Toronto, Ontario M5S 3H4, Canada. ¹²Department of Physics, National Sun Yat-Sen University No. 70, Lianhai Road, Gushan District, Kaohsiung City 804, Taiwan. ¹³Department of Physics and Astronomy, West Virginia University, Morgantown, West Virginia 26506, USA. ¹⁴Canadian Institute for Theoretical Astrophysics, 60 St George Street, Toronto, Ontario M5S 3H8, Canada. ¹⁵Perimeter Institute, 31 Caroline Street, Waterloo N2L 2Y5, Canada. ¹⁶Indian Institute of Science Education and Research Mohali, Knowledge City, Sector 81, SAS Nagar, Manauli, PO 140306, India.

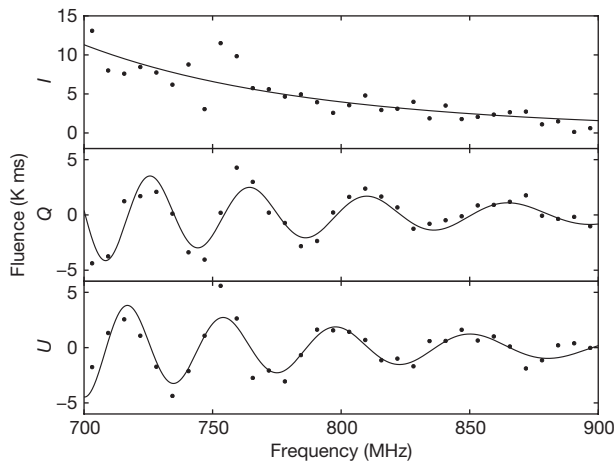


Figure 2 | FRB 110523 spectra in total intensity and polarization. Plotted is the pulse fluence (time-integrated flux) for total intensity (Stokes I), and linear polarization (Stokes Q and U). Solid curves are model fits. In addition to noise, scatter in the measurement around the models is due to the scintillation visible in Fig. 1. The decline of intensity with frequency is primarily due to motion of the telescope beam across the sky and is not intrinsic to the source.

By fitting a model to the burst data, we found the detection significance to be over 40σ , with fluence $3.79(15)$ K ms at our centre spectral frequency of 800 MHz. The burst has a steep spectral index of $-7.8(4)$, which we attribute to telescope motion. The peak antenna temperature at 800 MHz is $1.16(5)$ K. We do not know the location of the source within the GBT beam profile, but if the source location were at beam centre where the antenna gain is 2 K Jy^{-1} the measured antenna temperature would translate to 0.6 Jy . Off-centre the antenna has lower gain so, as in previously reported bursts, this is a lower limit to the FRB flux. The intrinsic full-width at half-maximum (FWHM) duration of the burst (with scattering removed) is $1.74(17)$ ms, similar to the widths of previously reported FRBs.

Allowing the dispersion relation to vary in the model, we found that the dispersive delay is proportional to $t_{\text{delay}} \approx \nu^{-1.998(3)}$, in close agreement to the expected ν^{-2} dependence for a cold, diffuse plasma. Following Katz²², this provides an upper limit on the density of electrons in the dispersing plasma of $n_e < 1.3 \times 10^7 \text{ cm}^{-3}$ at 95% confidence and a lower limit on the size of the dispersive region of $d > 10$ astronomical units (AU). This limit improves upon limits from previous bursts^{22–24} and rules out a flare star as the source of FRB 110523, because stellar coronas are denser and less extended by at least an order of magnitude²⁵. Flare stars are the last viable Galactic-origin model for FRB sources, so we take the source to be extragalactic.

We found strong linear polarization, with linearly polarized fraction $44(3)\%$. Linearly polarized radio sources exhibit Faraday rotation of the polarization angle on the sky by angle $\varphi_{\text{Far}} = \text{RM} \times \lambda^2$, where λ is the wavelength and the rotation measure RM (a measure of magnetization) is the line-of-sight component of the magnetic field weighted by the electron density:

$$\text{RM} = 0.812 \text{ rad m}^{-2} \int \frac{n_e}{\text{cm}^{-3}} \frac{B_{\parallel}}{\mu\text{G}} \frac{dl}{\text{pc}}$$

We detected the expected λ^2 modulation pattern in the polarization as shown in Fig. 2. The best-fitting RM is $-186.1(1.4) \text{ rad m}^{-2}$. All radio telescopes have polarization leakage, an instrument-induced false polarization of unpolarized sources. We mapped the leakage at GBT across the beam profile and throughout the passband and found that leakage can produce false linear polarization as large as 10% and false circular polarization as large as 30%. Leakage-produced apparent polarization lacks the λ^2 wavelength dependence that we see in the linear polarization data and cannot produce the 44% polarization we

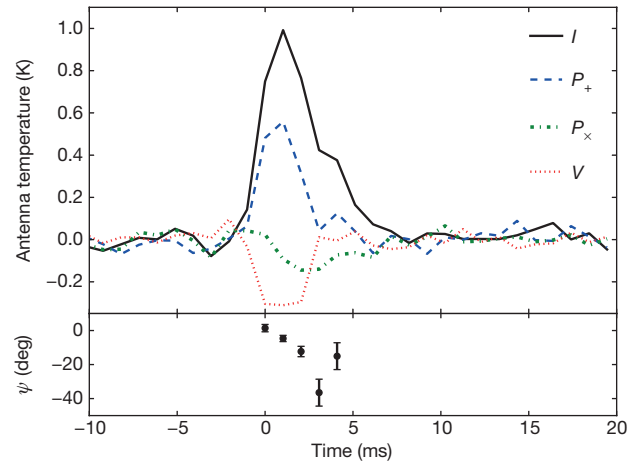


Figure 3 | Polarized pulse profiles averaged over spectral frequency. Plotted is total intensity (I), linear polarization (P_+ and P_-), and circular polarization (V , which may be instrumental). Before taking the noise-weighted mean over frequency, the data are scaled to 800 MHz using the best-fit spectral index and the linear polarization is rotated to compensate for the best-fit Faraday rotation. The linear polarization basis coordinates are aligned with (+), and rotated with respect to (\times), the mean polarization over time. The bottom panel shows the polarization angle (where measurable) in these coordinates. The error bars show the standard deviation of 10,000 simulated measurements with independent noise realizations.

detect so we conclude that the linear polarization is of astronomical origin rather than due to leakage.

The rotation measure and dispersion measure we detected imply an electron-weighted average line-of-sight component of the magnetic field of $0.38 \mu\text{G}$, compared to typical large-scale fields of $\sim 10 \mu\text{G}$ in spiral galaxies²⁶. This field strength is a lower limit for the magnetized region owing to cancellations along the line of sight. Also, the magnetized region may only weakly overlap the dispersing region and so electron weighting may not be representative.

The magnetization we detected is probably local to the FRB source rather than in the Milky Way or the intergalactic medium. Models of Faraday rotation within the Milky Way predict a contribution of $\text{RM} = 18(3) \text{ rad m}^{-2}$ for this line of sight, while the intergalactic medium can contribute at most $|\text{RM}| = 6 \text{ rad m}^{-2}$ on a typical line of sight from this redshift²⁷.

We detected a rotation of the polarization angle over the pulse duration of $-0.25(5) \text{ rad ms}^{-1}$, shown in Fig. 3. Such rotation of polarization is often seen in pulsars and is attributed to the changes in the projection of the magnetic field compared to the line of sight as the neutron star rotates²⁸.

We also detected circular polarization at roughly the 23% level, but that level of polarization might be due to instrumental leakage. Faraday rotation is undetectable for circular polarization, so the λ^2 modulation we use to identify astronomical linear polarization is not available as a tool to rule out leakage. For these reasons we do not have confidence that the detected circular polarization is of astronomical origin.

Radio emissions are often scattered: lensing by plasma inhomogeneities creates multiple propagation paths, with individual delays. We observed two distinct scattering timescales in the FRB 110523 data, indicating the presence of two scattering screens. In five previous FRB detections an exponential tail in the pulse profile was detected, interpreted as the superposition of delayed versions of the narrower intrinsic profile. The average scattering time constant for FRB 110523 is $1.66(14)$ ms at 800 MHz, with the expected decrease with spectral frequency as shown in Extended Data Fig. 2. We also detected scintillation, the variation of intensity with frequency due to multi-path interference. We measured a scintillation de-correlation bandwidth of $f_{\text{dc}} = 1.2(4) \text{ MHz}$ (see Extended Data Fig. 3), indicating a second source

of scattering with delays of the order of $1/f_{\text{dc}} \approx 1 \mu\text{s}$. This scintillation is consistent with Galactic expectations for this line of sight.

Scintillation should occur only if the first scattering screen is unresolved by the second, and we used this fact to constrain the bulk of the scattering material in the first screen to lie within 44 kpc of the source—roughly the scale of a galaxy (see Methods). It was previously unknown whether the approximately millisecond scattering observed in FRBs was due to weakly scattering material broadly distributed along the line of sight or strong scattering near the source²⁹, but our detection of scintillation eliminates the distributed scattering models. The observed scattering is too strong to be caused by the disks of host galaxies²² and therefore the FRB source must be associated with either a strongly scattering compact nebula or with the dense inner regions of its host galaxy. Either could produce the observed rotation measure, whereas most lines of sight through the interstellar medium of a randomly oriented galactic disk produce a rotation measure an order of magnitude smaller than we observed²⁷.

Magnetization and scattering located near the FRB source disfavour models that involve collisions of compact objects such as white dwarfs or neutron stars¹³ since these older stellar populations are generally not associated with compact nebulae, nor are they preferentially found near galactic centres. In contrast, a variety of models involving young stellar populations—including magnetar starquakes, delayed formation of black holes after core-collapse supernovae, and pulsar giant pulses^{15–17}—provide natural explanations for the properties we observe. Here scattering and magnetization occur in the surrounding young supernova remnants or star-forming regions, and the polarization properties we report are plausible given that these proposed emission mechanisms involve spinning magnetized compact objects. Precise model testing, beyond these general comments, will have to wait for more data which will determine whether the magnetization and scattering features we report are generic.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 July; accepted 25 September 2015.

Published online 2 December 2015.

- Lorimer, D. R., Bailes, M., McLaughlin, M. A., Narkevic, D. J. & Crawford, F. A bright millisecond radio burst of extragalactic origin. *Science* **318**, 777–780 (2007).
- Keane, E. F., Stappers, B. W., Kramer, M. & Lyne, A. G. On the origin of a highly dispersed coherent radio burst. *Mon. Not. R. Astron. Soc.* **425**, L71–L75 (2012).
- Thornton, D. *et al.* A population of fast radio bursts at cosmological distances. *Science* **341**, 53–56 (2013).
- Spitler, L. G. *et al.* Fast radio burst discovered in the Arecibo pulsar ALFA survey. *Astrophys. J.* **790**, 101 (2014).
- Burke-Spolaor, S. & Bannister, K. W. The Galactic position dependence of fast radio bursts and the discovery of FRB011025. *Astrophys. J.* **792**, 19 (2014).
- Petroff, E. *et al.* A real-time fast radio burst: polarization detection and multiwavelength follow-up. *Mon. Not. R. Astron. Soc.* **447**, 246–255 (2015).
- Ravi, V., Shannon, R. M. & Jameson, A. A fast radio burst in the direction of the Carina dwarf spheroidal galaxy. *Astrophys. J. Lett.* **799**, L5 (2015).
- Rane, A. *et al.* A search for rotating radio transients and fast radio bursts in the Parkes high-latitude pulsar survey. *Mon. Not. R. Astron. Soc.* <http://dx.doi.org/10.1093/mnras/stv2404> (2015); preprint at <http://arxiv.org/abs/1505.00834>.
- Taylor, M. *et al.* The core collapse supernova rate from the SDSS-II supernova survey. *Astrophys. J.* **792**, 135 (2014).
- Inoue, S. Probing the cosmic reionization history and local environment of gamma-ray bursts through radio dispersion. *Mon. Not. R. Astron. Soc.* **348**, 999–1008 (2004).
- Ioka, K. The cosmic dispersion measure from gamma-ray burst afterglows: probing the reionization history and the burst environment. *Astrophys. J. Lett.* **598**, L79–L82 (2003).
- Loeb, A., Shvartzvald, Y. & Maoz, D. Fast radio bursts may originate from nearby flaring stars. *Mon. Not. R. Astron. Soc.* **439**, L46–L50 (2014).
- Kulkarni, S. R., Ofek, E. O., Neill, J. D., Zheng, Z. & Juric, M. Giant sparks at cosmological distances? *Astrophys. J.* **797**, 70 (2014).
- Geng, J. J. & Huang, Y. F. Fast radio bursts: collisions between neutron stars and asteroids/comets. *Astrophys. J.* **809**, 24 (2015).
- Lyubarsky, Y. A model for fast extragalactic radio bursts. *Mon. Not. R. Astron. Soc.* **442**, L9–L13 (2014).
- Falcke, H. & Rezzolla, L. Fast radio bursts: the last sign of supramassive neutron stars. *Astron. Astrophys.* **562**, A137 (2014).
- Connor, L., Sievers, J. & Pen, U.-L. Non-cosmological FRBs from young supernova remnant pulsars. Preprint at <http://arxiv.org/abs/1505.05535> (2015).
- Chang, T.-C., Pen, U.-L., Bandura, K. & Peterson, J. B. An intensity map of hydrogen 21-cm emission at redshift $z \approx 0.8$. *Nature* **466**, 463–465 (2010).
- Masui, K. W. *et al.* Measurement of 21 cm brightness fluctuations at $z \sim 0.8$ in cross-correlation. *Astrophys. J. Lett.* **763**, L20 (2013).
- Switzer, E. R. *et al.* Determination of $z \approx 0.8$ neutral hydrogen fluctuations using the 21-cm intensity mapping autocorrelation. *Mon. Not. R. Astron. Soc.* **434**, L46–L50 (2013).
- Cordes, J. M. & Lazio, T. J. W. NE2001.I. A new model for the galactic distribution of free electrons and its fluctuations. Preprint at <http://arxiv.org/abs/astro-ph/0207156> (2002).
- Katz, J. I. Inferences from the distributions of fast radio burst pulse widths, dispersion measures and fluences. Preprint at <http://arxiv.org/abs/1505.06220> (2015).
- Tuntsov, A. V. Dense plasma dispersion of fast radio bursts. *Mon. Not. R. Astron. Soc.* **441**, L26–L30 (2014).
- Dennison, B. Fast radio bursts: constraints on the dispersing medium. *Mon. Not. R. Astron. Soc.* **443**, L11–L14 (2014).
- Maoz, D. *et al.* Fast radio bursts: the observational case for a Galactic origin. *Mon. Not. R. Astron. Soc.* **454**, 2183–2189 (2015).
- Widrow, L. M. Origin of galactic and extragalactic magnetic fields. *Rev. Mod. Phys.* **74**, 775–823 (2002).
- Oppermann, N. *et al.* Estimating extragalactic Faraday rotation. *Astron. Astrophys.* **575**, A118 (2015).
- Radhakrishnan, V. & Cooke, D. J. Magnetic poles and the polarization structure of pulsar radiation. *Astrophys. Lett.* **3**, 225–229 (1969).
- Macquart, J.-P. & Koay, J. Y. Temporal smearing of transient radio sources by the inter-galactic medium. *Astrophys. J.* **776**, 125 (2013).

Acknowledgements K.M. is supported by the CIFAR Global Scholars Program. T.-C.C. acknowledges support from MOST grant 103-2112-M-001-002-MY3. X.C. and Y.-C.L. are supported by MOST 863 programme 2012AA121701, CAS XDB09000000 and NSFC 11373030. P.T.T. acknowledges support from NSF Award 1211781. J.B.P. acknowledges support from NSF Award 1211777. Computations were performed on the General Purpose Cluster supercomputer at the SciNet HPC Consortium.

Author Contributions K.M. integrated the FRB search routines into a software program; calibrated and filtered the raw FRB event data; performed scintillation analysis; led survey planning; produced Figs 2 and 3 and Extended Data Fig. 3; and contributed to model fits to the FRB event, result interpretation, beam characterization, observations, data handling, and data validation. H.-H.L. performed the visual search of the search of over 6,000 images, and discovered the FRB event. H.-H.L. also coproduced Fig. 1, produced Extended Data Fig. 2, and contributed to the FRB search software program, observations, data handling, and data validation. J.S. wrote dedispersion and FRB search software routines; performed model fits to the FRB event (including extracting the dispersion measure, rotation measure, scattering tail, and polarization angle swing); and contributed to result interpretation. C.J.A. contributed to observations, data handling, and data validation. T.-C.C. contributed to survey planning, observations, and data validation. X.C. contributed to data validation. A.G. contributed to FRB search algorithm validation. M.J. contributed to observations and data validation. C.-Y.K. contributed to observations and data validation. Y.-C.L. performed scintillation analysis on the foreground pulsar and contributed to data validation. Y.-W.L. contributed to polarization leakage characterization, calibration methods, and data validation. M.McL. contributed to result interpretation, analysis of the follow-up data, scintillation analysis on the foreground pulsar, and edited the manuscript. U.-L.P. carried out Faraday rotation measure synthesis (resulting in the detection of linear polarization) and contributed to result interpretation, scintillation analysis, survey planning, and data validation. J.B.P. led manuscript preparation and contributed to result interpretation, survey planning and data validation. A.R. surveyed archival multi-wavelength catalogues for coincident sources, coproduced Fig. 1, produced Extended Data Fig. 3, and added event simulation functionality to the FRB search software program. P.T.T. contributed to observations and data validation and editing of the manuscript. T.V. led the observational campaign and contributed to calibration methods, survey planning, data handling, and data validation. J.K.Y. contributed to data validation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.M. (kiyo@physics.ubc.ca).

METHODS

Data and pre-processing. Our survey was conducted with the GBT linearly polarized prime-focus 800-MHz receiver. For the digital back-end we used the GBT Ultimate Pulsar Processing Instrument. The data were collected with the telescope scanning 4° per minute at constant altitude angle.

To act as a stable flux reference, a broadband noise source injects power at the feed point, producing a square wave of intensity with period locked at 64 times the 1.024-ms cadence. In the on-state the noise source increases the total power by approximately 10%. The switching noise source must be removed from the data before the search for transients can proceed. This is done by accumulating, over the one-minute scan, the periodic component with a period of 64 ms. The data are then normalized to the noise source amplitude in each spectral channel, providing an approximate bandpass calibration, and the noise source waveform is subtracted. For the search phase this level of calibration is sufficient and no absolute calibration is applied.

Analysis of the discovered event requires a more rigorous calibration than the search. We separately reference the vertical and horizontal polarization signals to the calibration noise source, with the noise source in turn referenced to a bright unpolarized point source (3C48) scanned 6.5 h before the event, providing an on-sky calibration at each frequency and polarization. This results in an overall absolute flux calibration uncertainty of 9% (ref. 19). To calibrate the phase of the cross-correlation between the two antenna polarizations, which we need to measure polarization parameters Stokes U and V , we assume that the noise source injects the same signal into each with the same phase. Laboratory tests of the 800-MHz receiver verify this assumption except in the two spectral resonances of the receiver and in the edges of the band, which we discard. This procedure produces a 1% polarization calibration at the centre of the beam. The polarization characteristics off beam centre are described below.

The data contain several spectral channels that are irrecoverably contaminated by man-made radio frequency interference, largely due to cell (mobile) phones. These are identified by anomalously high variance or skewness relative to other channels and all data from these channels are discarded. A total of 3,836 out of 4,096 channels (93.6%) pass the radio-frequency interference cuts.

Prior to searching the data for FRBs, Galactic and extragalactic synchrotron continuum emission is removed. Such emission is broadband and varies on much longer timescales than FRB events and can thus be removed by a variety of algorithms. For the search, where computational efficiency is a concern, a continuum template is formed for each 38-s block of data by performing a mean over frequency. This template is then correlated against each spectral channel and the contribution subtracted.

When analysing the discovered event, computational complexity is less of a concern so we high-pass filter the data on 200-ms timescales. The filtering substantially reduces the variance of the data, and we perform another iteration of identification of spectral channels contaminated with radio-frequency interference.

Searching the data. To search for FRBs we concentrate the energy of possible events into a few pixels of an image, using a dispersion-removal algorithm we developed. In the array of spectra shown in Extended Data Fig. 1, the event is spread out in both time and frequency. We need to remove this dispersion, aligning the arrival times across frequency, then average over frequency to produce a time series that has the pulse energy localized. Since we do not know the dispersion measure a priori we dedisperse over a range of trial values of DM of 0–2,000 pc cm⁻³. The dedispersion process produces a set of frequency-averaged intensities versus time and DM. We use a modified tree dedispersion algorithm³⁰. We developed a recursive program for this algorithm that, running on a single desktop computer, carries out the dispersion removal in 10% of real time.

After transforming to DM–time space, we need to search each DM for bursts of unknown duration and unknown profile, which we accomplish using a set of boxcar integrals over time, of lengths ranging from 1 ms to the block length of 38 s. Blocks overlap by 8 s so that events straddling blocks are not missed. The search algorithm also accumulates noise statistics at each DM for each boxcar length. The procedure is easily parallelized by distributing data files among nodes of a large computer array. A software package used to search our data for transient events is publicly available: https://github.com/kiyo-masui/burst_search.

The above procedure produced 6,496 DM–time plots, which we visually inspected. We find only one clear FRB candidate—FRB 110523—but the search also turned up the previously known pulsars J2139+00 and J0051+0423, roughly in line with expectations given the survey parameters.

We have yet to perform a detailed analysis of the completeness of our search but taken at face value our single detection implies an all-sky rate of about 5×10^3 per day above a fluence of threshold of ~ 1 Jy ms, assuming an effective sky coverage of 0.3 square degrees, which is in line with previous estimates.

To provide a set of training templates for the visual search, simulated rectangular pulses were added onto a sample of data which included no significant events.

An example of a simulated event is shown in Extended Data Fig. 1. The simulated event shows a characteristic ‘hourglass’ feature in the DM–time plots.

Modelling the pulse profile and polarization. We use Markov Chain Monte Carlo methods to fit a model to the FRB event and measure its properties. Throughout the analysis we assume the noise is Gaussian and treat it as uncorrelated between channels, with per-channel weights estimated from their variances. This simplification allows us to forgo the time-consuming process of Fisher matrix statistical analysis. The assumption is slightly incorrect: the data are χ^2 distributed with 50 degrees of freedom. We also find that adjacent frequency channels are actually 2.5% correlated by the Fourier transform filter used for spectral channelization. No significant correlation is detected between more widely separated channels. We account for adjacent channel correlation by increasing all errors by 2.5%.

To create a model intensity profile for comparison to the data we begin with a Gaussian pulse profile in time, with width σ , which is independent of frequency. This is convolved with a one-sided exponential scattering kernel with a frequency-dependent duration to yield the normalized pulse profile:

$$f(\nu, t) = \left[\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{t^2}{2\sigma^2}\right) \right] \otimes \left[\theta(t) \frac{1}{\tau_\nu} \exp\left(-\frac{t}{\tau_\nu}\right) \right] \quad (1)$$

where $\theta(t)$ is the Heaviside step function, $\tau_\nu = \tau(\nu/\nu_{\text{ref}})^{-4}$ is the frequency dependence expected for scattering, and τ is the scattering time at reference frequency ν_{ref} . In the final spectrum we allow for spectral index α of the overall intensity and delay the pulse for dispersion:

$$I(\nu, t) = A \left(\frac{\nu}{\nu_{\text{ref}}} \right)^\alpha f(\nu, t - t_\nu) \quad (2)$$

where A is the burst amplitude at ν_{ref} , $t_\nu = t_0 + \text{DM} \times \text{DM}_0(\nu^{-2} - \nu_{\text{ref}}^{-2})$, $\text{DM}_0 = 4,148.808 \text{ MHz}^2 \text{ pc}^{-1} \text{ cm}^3 \text{ s}$, DM is the dispersion measure of the burst, and t_0 is the burst arrival time at ν_{ref} . While in principle the choice of the reference frequency ν_{ref} is arbitrary, in practice we find a value of 764.2 MHz, near the centre of the signal-to-noise weighted band, substantially decorrelates the fit parameters. This constitutes our base unpolarized model. Circular polarization is modelled in the same way as total intensity.

Our base linearly polarized model is the same as the unpolarized model with an added Faraday rotation factor:

$$[Q + iU](\nu, t) = I \exp[2i\text{RM}(\lambda^2 - \lambda_{\text{ref}}^2) + i\varphi_0] \quad (3)$$

where RM is the rotation measure, φ_0 is the polarization angle at the reference frequency and pulse centre, and I is the model for intensity given above. We find the likelihood surfaces are quite close to Gaussian, and so the Markov chains converge quickly. We run an initial short chain to estimate the covariance matrix, then run four chains of length 500,000 steps to estimate parameters. This approach gives good convergence (the Gelman–Rubin convergence $r - 1$ is typically 0.005).

To search for time dependence of the polarization angle, we extend the model to allow the polarization angle to vary linearly with time. We performed this fit two ways: (1) apply the phase gradient to the pre-scattering Gaussian burst and then convolve with the scattering kernel, and (2) apply the gradient to the scattering-convolved burst profile. While the first is more physically natural if the rotation happens at the burst source before scattering, we find that the second (post-scattering gradient) provides a significantly better fit (5.4σ significance compared to 2.1σ) and quote results for this fit. We attribute the higher significance of the less physical model to substructure in the polarized pulse that is poorly modelled by a Gaussian with linearly changing polarization angle. We do not have a high enough signal-to-noise ratio to further investigate the substructure but the conclusion that the polarization angle rotates is robust.

The dispersion delay as a function of frequency is expected to follow a ν^{-2} power law, scattering time should have frequency dependence near ν^{-4} , and the Faraday rotation angle should be proportional to ν^{-2} . We extend the model used in the Markov chains to test these predictions, fitting for the dispersion measure index, scattering index, and Faraday rotation index. All fit parameters are listed in Extended Data Table 1 with results grouped by independent fits.

To check our analysis software and calibration (especially the polarization sign) we performed observations of pulsar B2319+60. The pulsar data were processed using the FRB pipeline and the Faraday rotation measured from a single pulse. The rotation measure was determined to be $-239.9(4) \text{ rad m}^{-2}$, in good agreement with the published value³¹, and under this sign convention the RM of the FRB is negative.

GBT beam. During the 2-s period over which the FRB pulse traverses the bandpass, the pointing centre of the GBT beam scans 8 arcmin, which is about half the FWHM beam width. The pulse intensity increases steeply during the arrival period, probably indicating that the source coordinates moved from the edge of the GBT

beam at the start of the arrival period to a position closer to the beam centre as lower frequencies arrived. The GBT beam is also wider at lower frequencies, which also contributes to the steep spectral index. Simulations indicate that this picture is consistent although, because the intrinsic spectral index of the source and the impact parameter of the scan relative to the source are unknown, we are unable to use this information to obtain a precise localization.

It is highly unlikely that the burst entered the telescope through a sidelobe. Because of its off-axis design, GBT has low sensitivity in its sidelobes. Simulated models of the 800-MHz receiver beam show the first sidelobe to be a ring around the primary beam with radius 0.6° , width 0.1° , and 30 dB less sensitivity than beam centre (boresight) (S. Srikanth, personal communication, November 2012). The second and third sidelobes have similar geometry, occurring 0.8° and 1.0° from boresight and suppressed by 37 dB and 40 dB, respectively. These near sidelobes do not cover much more sky area than the main beam and with their dramatically lower sensitivity it is unlikely that the lobe would contribute to the burst detection rate.

Subsequent sidelobes have even less sensitivity but cover more area. They are ruled out by the observed spectrum of FRB 110523. The radial width of the sidelobes is 0.1° and their radial locations are inversely proportional to observing frequency. As such, if the burst had entered a far sidelobe we would have observed far more spectral structure, with several peaks and nulls. Even the previously discussed first sidelobe is in tension with the observed spectrum when accounting for the added spectral structure expected from about 0.1° of scanning during the pulse arrival period. For the first three sidelobes it is possible that, though an improbable coincidence, the telescope's scanning could cancel the location spectral dependence of the sidelobes. However, as previously argued, a source location in the sidelobes is unlikely owing to their combination of low sensitivity and low area.

To determine the polarization properties of the primary beam, we performed on-axis and off-axis measurements of the beam using both bright point sources and pulsars. Such measurements are crucial for our survey's primary science goal of mapping cosmic structure through the 21-cm line. We find that although GBT's off-axis design reduces sidelobe amplitude it leads to substantial polarization leakage in the primary beam. On boresight, the leakage from total intensity to polarization is less than 1%. Off boresight, leakage peaks at approximately 0.2° in the azimuth direction. Leakage from Stokes I to Q/U is several per cent of the forward gain and from Stokes I to V it is as high as 10%. When comparing to the gain at that location in the beam instead of the forward gain, these numbers translate to 10% leakage to linear polarization and 30% leakage to circular. The leakage is only weakly dependent on frequency. These measurements are in agreement with simulated beam models.

The observed polarization angle rotation over the duration of the pulse cannot be due to leakage. The rotation occurs in each frequency bin over a few milliseconds, during which time the GBT beam centre moves just 7 milliarcseconds. Gradients of the leakage pattern at such small difference of angle are much too small to explain the change of polarization angle. To achieve a signal-to-noise ratio that is sufficient to detect the angle swing it is necessary to integrate over frequency, introducing the 2-s timescale associated with dispersion delay, but the integrand is composed of millisecond differences of polarization angle, making the 2-s timescale irrelevant.

Scintillation. Since we see the FRB pulse only for a few milliseconds we have no information on the variation of the flux on longer timescales, and concentrate on quantifying the scintillation-induced variation of intensity with frequency by calculating the de-correlation bandwidth. We first form $\delta T(\nu) \equiv T(\nu)/T_{\text{smooth}}(\nu)$ where $T_{\text{smooth}}(\nu)$ is the power-law fit to the spectrum, accounting for the intrinsic spectrum of the event as well as the frequency dependence and motion of the telescope beam. We then form the correlation function:

$$\xi(\Delta\nu) \equiv \langle \delta T(\nu) \delta T(\nu + \Delta\nu) \rangle_\nu \quad (4)$$

This correlation function is estimated from the observed spectrum and is shown in Extended Data Fig. 3.

To estimate the de-correlation bandwidth, f_{dc} , from the observed correlation function, we fit it to the Fourier transform of an exponential scattering function³²:

$$\xi_{\text{model}}(\Delta\nu) = \frac{m}{f_{\text{dc}}^2 + \Delta\nu^2} \quad (5)$$

This fit yields $f_{\text{dc}} = 1.2(4)$ MHz and $m = 0.26(8)$. The errors on the measurement of the correlation function depend on the underlying statistics of the scintillation, which are both non-Gaussian and model-dependent³³. We estimate the errors in Extended Data Fig. 3 through simulations, with errors on fit parameters subsequently expanded to account for modelling uncertainties.

Two-screen model for scintillation and scattering. The observed scintillation de-correlation bandwidth is comparable to that observed for the Galactic

pulsar J2139+00, less than two degrees away from FRB 110523 on the sky and at a distance of 3 kpc based on its dispersion measure³⁴, indicating that the scintillation arises from the Galactic interstellar medium.

A familiar form of scintillation in optical astronomy is the twinkling of stars. Optical scintillation is due to turbulence in the atmosphere and is commonly modelled by projecting the optical medium onto a screen above the telescope with micro-images appearing in the plane of this screen. For stars, a rapid variation of flux with frequency is seen because stars have angular size small enough that light emitted from opposite edges of the stellar disk has path length difference less than a wavelength. Stars are said to be unresolved by the scintillation screen, meaning that they are indistinguishable from point sources. The multipath interference changes with time because of turbulent motions in the atmosphere. Planets, in contrast to stars, have angular size resolved by the screen, so the flux variations are typically a small fraction of the total flux. For similar reasons, among radio sources, pulsars often show scintillation, while the much larger extragalactic radio sources do not. At radio wavelengths scattering occurs in the intervening plasma rather than the atmosphere.

To model scintillation and scattering for FRB 110523 we project the intervening material onto two screens, representing the material in the Milky Way and in the host galaxy, respectively. We use two screens because the scintillation and scattering have very different timescales, which precludes modelling with a single screen. As with optical scintillation each screen produces a halo of micro-images, which can be considered scattering sites. Propagation via a micro-image at the edge of a halo requires a longer propagation time from source to observer than micro-images near the centre. In our model the delays associated with the Galactic screen produce the microsecond scintillation path differences, while the host screen path differences produce the 1.6-ms exponential tail of the pulse profile.

In our two-screen model the presence of strong scintillation indicates that the host screen is unresolved by the Galactic screen, and this allows an estimate of the host screen position. We assume that the position of the Galactic screen is the characteristic thickness of the ionized Galactic plane $D = 1$ kpc. The angular size of the Galactic screen is then given by $\theta = \sqrt{2c\tau/D} \approx 1$ milliarcsecond and the resolving power of the Galactic screen is $\rho = \lambda/(\theta D) \approx 600$ nanoarcseconds. The scintillation would be washed out if the host screen exceeded this angular size. This small angular size, combined with the 1.6-ms scattering time, places the host scattering screen within 44 kpc of the source, assuming the maximum source distance of about 1 Gpc (constrained by the observed dispersion measure).

To further test our scintillation and scattering model, we compared the scintillation of the main pulse to the scintillation in the scattering tail by cross-correlating the intensity spectrum early in the pulse to the spectrum late in the pulse. To obtain the early pulse spectrum, we use a filter matched to the Gaussian part of the profile with no scattering tail. For the late part we use a filter matched to the tail beginning 3 ms into the pulse. The cross de-correlation bandwidth is $f_{\text{dc}} = 1.3(5)$ MHz, compared to $f_{\text{dc}} = 1.1(6)$ MHz and $f_{\text{dc}} = 1.0(4)$ MHz for the early and late pulses, respectively. Correlation amplitudes are $m = 0.30(9)$, $m = 0.18(8)$, and $m = 0.47(13)$ for the cross-correlation, early pulse, and late pulse, respectively. These are all consistent with the level of scintillation measured for the full pulse, indicating that the most direct path and scattering-delayed micro-images share a common scintillation-induced spectrum. The scintillation source is therefore separate from the source of the scattering tail, and we place them in the Milky Way and host galaxy respectively.

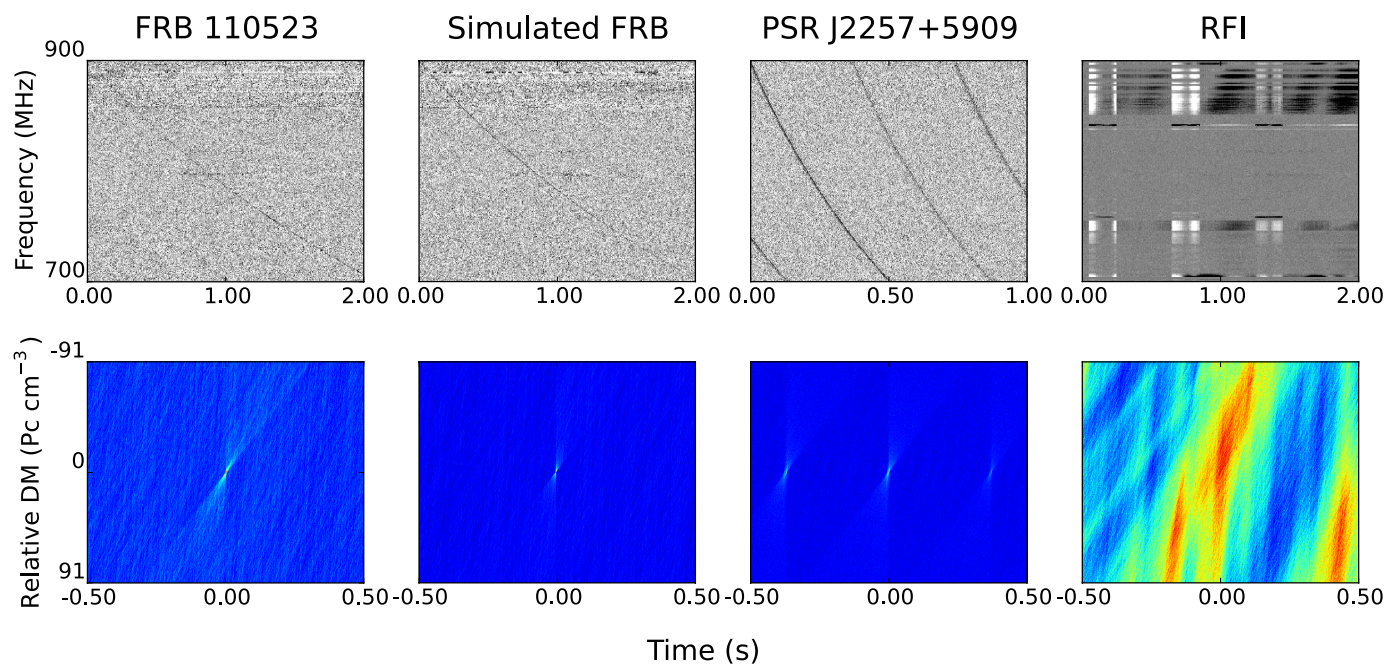
Follow-up observations. We carried out observations at the position of FRB 110523 from 700 MHz to 900 MHz at three separate epochs on MJD 57134, MJD 57135, and MJD 57157 for durations of 1.8 h, 1.8 h, and 3 h, respectively. We detected no bursts with DMs in the range $0-5,000 \text{ pc cm}^{-3}$ with significance greater than 6σ . We also performed a periodicity search on the data, and detected no pulsar candidates. The estimated limiting flux density of this search, assuming a pulsar duty cycle of 10%, was 0.04 mJy .

Counterpart sources. To identify possible optical counterpart source candidates we searched the Sloan Digital Sky Survey (DR12) catalogues³⁵ throughout a region centred on the position of the radio beam at the time the pulse arrived at 700 MHz. The beam size of the GBT is 15 arcmin at FWHM, but we expanded the search area to a diameter of 30 arcmin to account for a source lying outside the FWHM beam area. Within this field there are 70 objects identified as galaxies in the catalogue, of which 40 are listed as having redshift less than 0.5. The 100% galactic completeness limit of SDSS photometry³⁶ is r-band magnitude 21. As such, all Milky-Way-like galaxies are included for $z < 0.28$, assuming an absolute magnitude of $M_r \approx -19.86$. No X-ray or gamma-ray sources are listed in the NASA/IPAC Extragalactic Database in this region.

Data availability. The raw data used in this publication are available at <http://www.cita.utoronto.ca/~kiyo/release/FRB110523>.

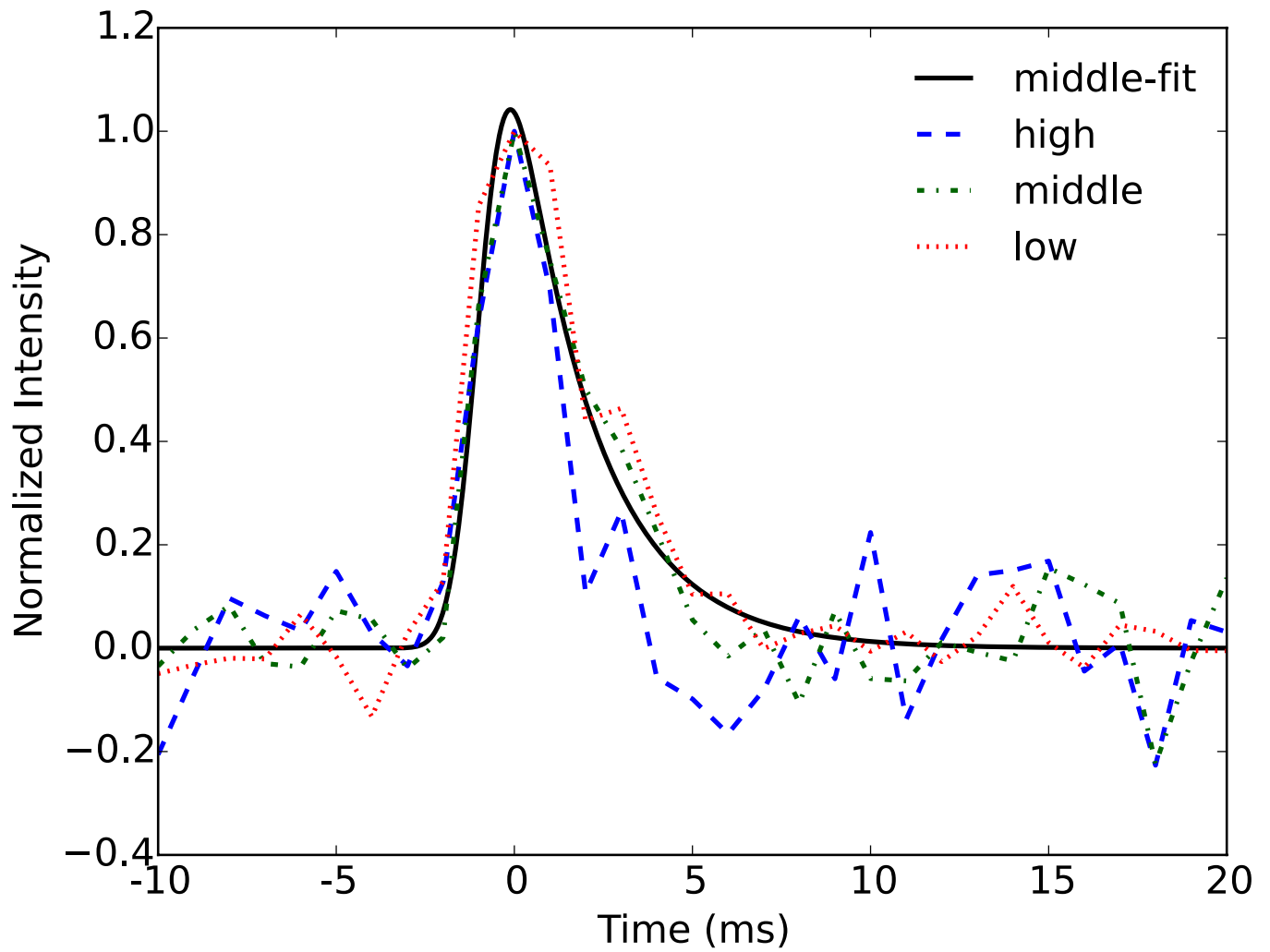
Code availability. The code used to search the data archive for FRB events is available at https://github.com/kiyo-masui/burst_search. The code used to analyse the discovered FRB is available at https://github.com/kiyo-masui/FRB110523_analysis.

30. Taylor, J. H. A sensitive method for detecting dispersed radio emission. *Astron. Astrophys. (Suppl.)* **15**, 367–369 (1974).
31. Hamilton, P. A. & Lyne, A. G. Faraday rotation measurements on 163 pulsars. *Mon. Not. R. Astron. Soc.* **224**, 1073–1081 (1987).
32. Lorimer, D. R. & Kramer, M. *Handbook of Pulsar Astronomy* Ch. 4 (Cambridge Univ. Press, 2004).
33. Johnson, M. D. & Gwinn, C. R. Ultra-high-resolution intensity statistics of a scintillating source. *Astrophys. J.* **755**, 179 (2012).
34. Taylor, J. H. & Cordes, J. M. Pulsar distances and the galactic distribution of free electrons. *Astrophys. J.* **411**, 674–684 (1993).
35. Alam, S. *et al.* The eleventh and twelfth data releases of the Sloan Digital Sky Survey: final data from SDSS-III. *Astrophys. J. Suppl.* **219**, 12 (2015).
36. Yasuda, N. *et al.* Galaxy number counts from the Sloan Digital Sky Survey commissioning data. *Astron. J.* **122**, 1104–1124 (2001).

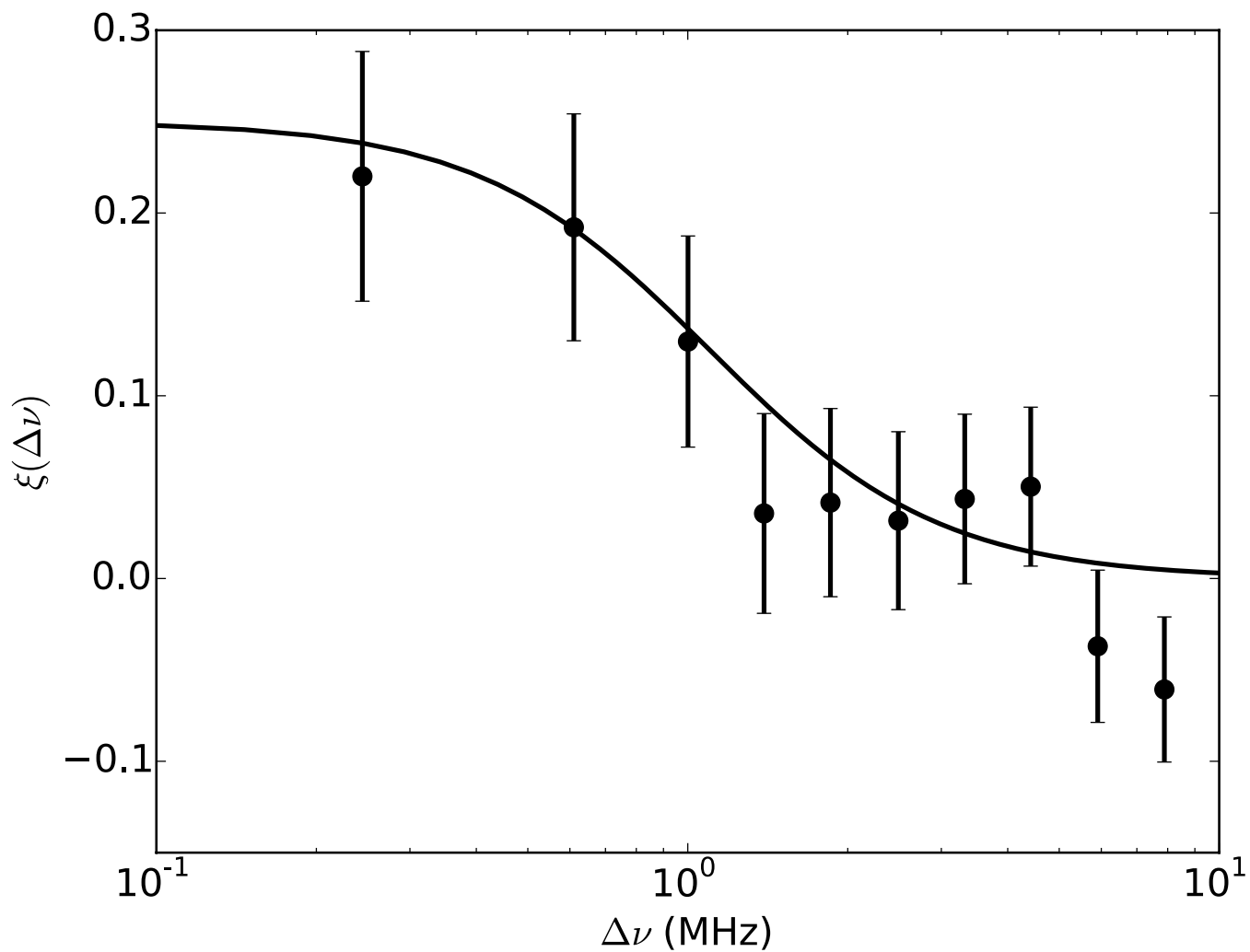


Extended Data Figure 1 | Events in frequency-time and DM-time space. From left to right are shown data for FRB 110523, a simulated FRB, a known pulsar PSR J2257+5909, and man-made radio frequency interference (RFI). Brightness temperature is shown in frequency-time space (upper panels) and the same data in DM-time space (lower panels). The relative dispersion measure is the difference between the DM

and the event DM; event DM values are 622.8 pc cm^{-3} , 610.3 pc cm^{-3} , 151.0 pc cm^{-3} and $1132.1 \text{ pc cm}^{-3}$ from left to right. The time axes of the frequency-time plots show time relative to the zero time in DM-time space. The colour scale in the lower panels represents broadband flux, with red showing a bright source.



Extended Data Figure 2 | Pulse profiles for FRB 110523 in three sub-bands. Each sub-band has width of 66 MHz. The pulse width decreases with frequency (at 2.6σ significance), consistent with models of scattering in the interstellar medium. Also shown in black is the best-fit model profile for the middle band.



Extended Data Figure 3 | Spectral brightness correlation function of FRB 110523. The intensity spectrum has structure that is correlated for frequency separations less than $f_{\text{dc}} = 1.2$ MHz. Error bars are the standard deviation of 3,268 simulated measurements with 817 independent noise realizations and are correlated.

Extended Data Table 1 | FRB 110523 parameters

Barycentric $\nu = \infty$ arrival (MJD)	55704.62939511
GBT boresight at 900 MHz arrival	RA = 21 ^h 45 ^m 31 ^s Dec = −00 ^d 15 ^m 23 ^s $l = 56.0795^\circ$ $b = -37.9435^\circ$
GBT boresight at 700 MHz arrival	RA = 21 ^h 45 ^m 12 ^s Dec = −00 ^d 09 ^m 37 ^s $l = 56.1215^\circ$ $b = -37.8234^\circ$
Dispersion measure (pc cm ^{−3})	623.30(6)
Fluence at 800 MHz (K ms)	3.79(15)
Spectral index	−7.8(4)
Unscattered pulse FWHM (ms)	1.73(17)
Scattering time at 800 MHz (ms)	1.66(14)
Linear polarization fraction (%)	44(3)
Rotation measure (rad m ^{−2})	−186.1(1.4)
Polarization rotation rate (rad ms ^{−1})	−0.25(5)
Dispersion measure index	−1.998(3)
Scattering index	−3.6(1.4)
Faraday rotation index	−1.7(2)

Arrival time and astrometric parameters as well as parameters for fits of the base-unpolarized, base-polarized, and extended models to antenna temperature data. The steep spectral index we observe is attributed to beam effects. Statistical uncertainties enclose the 68% confidence interval of the measurement. RA, right ascension; Dec, declination. MJD, modified Julian day. l , Galactic longitude; b , Galactic latitude.

A dynamic magnetic tension force as the cause of failed solar eruptions

Clayton E. Myers^{1,2}, Masaaki Yamada², Hantao Ji^{1,2,3}, Jongsoo Yoo², William Fox², Jonathan Jara-Almonte^{1,2}, Antonia Savcheva⁴ & Edward E. DeLuca⁴

Coronal mass ejections are solar eruptions driven by a sudden release of magnetic energy stored in the Sun's corona¹. In many cases, this magnetic energy is stored in long-lived, arched structures called magnetic flux ropes^{2–5}. When a flux rope destabilizes, it can either erupt and produce a coronal mass ejection or fail and collapse back towards the Sun^{6–8}. The prevailing belief is that the outcome of a given event is determined by a magnetohydrodynamic force imbalance called the torus instability^{9–14}. This belief is challenged, however, by observations indicating that torus-unstable flux ropes sometimes fail to erupt¹⁵. This contradiction has not yet been resolved because of a lack of coronal magnetic field measurements and the limitations of idealized numerical modelling. Here we report the results of a laboratory experiment¹⁶ that reveal a previously unknown eruption criterion below which torus-unstable flux ropes fail to erupt. We find that such 'failed torus' events occur when the guide magnetic field (that is, the ambient field that runs toroidally along the flux rope) is strong enough to prevent the flux rope from kinking. Under these conditions, the guide field interacts with electric currents in the flux rope to produce a dynamic toroidal field tension force that halts the eruption. This magnetic tension force is missing from existing eruption models, which is why such models cannot explain or predict failed torus events.

For a laboratory experiment to study ideal instability solar eruption mechanisms such as the torus instability, it must adhere to the standard storage-and-release model for solar eruptions. According to this model, eruptions are triggered by transient events in the corona rather than by dynamic changes at the solar surface¹. For an arched flux rope, the relative invariance of the solar surface translates to a slow driving requirement at the two 'line-tied' (anchored) footpoints. Previous laboratory arched flux rope experiments^{17–19} have deviated from the storage-and-release model by relying on the dynamic injection of either plasma or magnetic flux at the footpoints to produce an eruption. In contrast, the present experiments¹⁶ enforce a strict separation of timescales between the footpoint driving time, τ_D , and the dynamic Alfvén time, τ_A , such that the observed eruptions must be driven by storage-and-release mechanisms (see Methods and Extended Data Tables 1 and 2).

Flux ropes in the solar corona are most susceptible to two ideal magnetohydrodynamic instabilities: the torus instability^{9–14} and the kink instability^{20–24} (see Methods). At present, the torus instability is thought to be the primary driver of eruptions¹³, while the kink is believed to play a secondary part⁷. The onset criteria for these instabilities are inextricably linked to the ambient potential magnetic field (also known as the vacuum field) in which the flux rope is embedded. On the Sun, the potential field is produced by sources located beneath the solar surface, while in the laboratory it is produced by fixed magnetic field coils located outside the plasma (see Extended Data Fig. 1). In either case, the potential field can be decomposed into two orthogonal components: the strapping field, which runs perpendicular to the flux rope, and the

guide field, which runs toroidally along it (see Extended Data Fig. 2). The strapping field is central to the torus instability in that it produces the strapping force, which counters the upward-driving 'hoop' force and restrains the flux rope (see Methods). The guide field, on the other hand, is central to the kink instability in that it reduces the magnetic twist in the flux rope (see Methods).

More quantitatively, the critical parameter for the torus instability is the potential field decay index¹⁰, n , which characterizes the spatial decay of the potential field (a high n value indicates a steep spatial decay and hence torus instability; see Methods). Likewise, the critical parameter for the kink instability is the edge safety factor^{25–27}, q_a (where a is the edge minor radius of the flux rope), which characterizes the inverse magnetic twist in the flux rope (a low q_a value indicates a high twist and hence kink instability; see Methods). Our laboratory experiments facilitate the independent control of n and q_a , enabling us to systematically explore the torus versus kink instability parameter space and to identify the stability boundaries.

The n versus q_a parameter space is scanned in the experiment by independently modifying the magnitude and the vertical (z) profile of each potential field component. Figure 1 compares two representative flux rope plasmas with different potential field settings: the flux rope in Fig. 1c has high q_a and low n such that it is stable, while the flux rope in Fig. 1d has low q_a and high n such that it erupts violently and repeatedly towards the wall of the machine. These are just two examples from a comprehensive scan of n and q_a , the results of which are shown in Fig. 2. Four distinct parameter regimes are readily identified in the experimental data. Three of these (the stable, eruptive, and failed kink regimes) are consistent with the present understanding of the torus and kink instabilities. In particular, the kink instability appears below $q_a \approx 0.8$ but does not necessarily drive an eruption. Only when the decay index also exceeds the observed torus threshold ($n \approx 0.8$) does the failed kink regime give way to the eruptive regime (consistent with numerical simulations⁷). Interestingly, the observed torus threshold of $n \approx 0.8$ is substantially lower than the theoretical expectation of $n = 3/2$. This reduced threshold is consistent with the theory of the 'partial torus instability', which accounts for the effect of the line-tied geometry on the hoop force²⁸. The fourth instability regime identified in Fig. 2, which we call the 'failed torus' regime, contradicts the widely held notion that the torus criterion is a sufficient condition for eruption. In this regime, kink-stable flux ropes that exceed the torus threshold fail to erupt. This behaviour cannot be explained in terms of the hoop and strapping forces alone. Instead, a magnetic tension force related to the toroidal guide field plays a crucial part.

To examine more carefully the physics of the failed torus regime, magnetic field data from a characteristic failed torus event are shown in Fig. 3. The height–time evolution of this event (Fig. 3a) shows that the plasma initially rises before saturating and then rapidly collapsing downward. Clues as to why this occurs are found in spatial plots of the toroidal current density, J_T (Fig. 3d, see Extended Data Table 3 for

¹Department of Astrophysical Sciences, Princeton University, Princeton, New Jersey 08544, USA. ²Princeton Plasma Physics Laboratory, Princeton, New Jersey 08543, USA. ³Laboratory for Space Environment and Physical Sciences, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China. ⁴Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA.

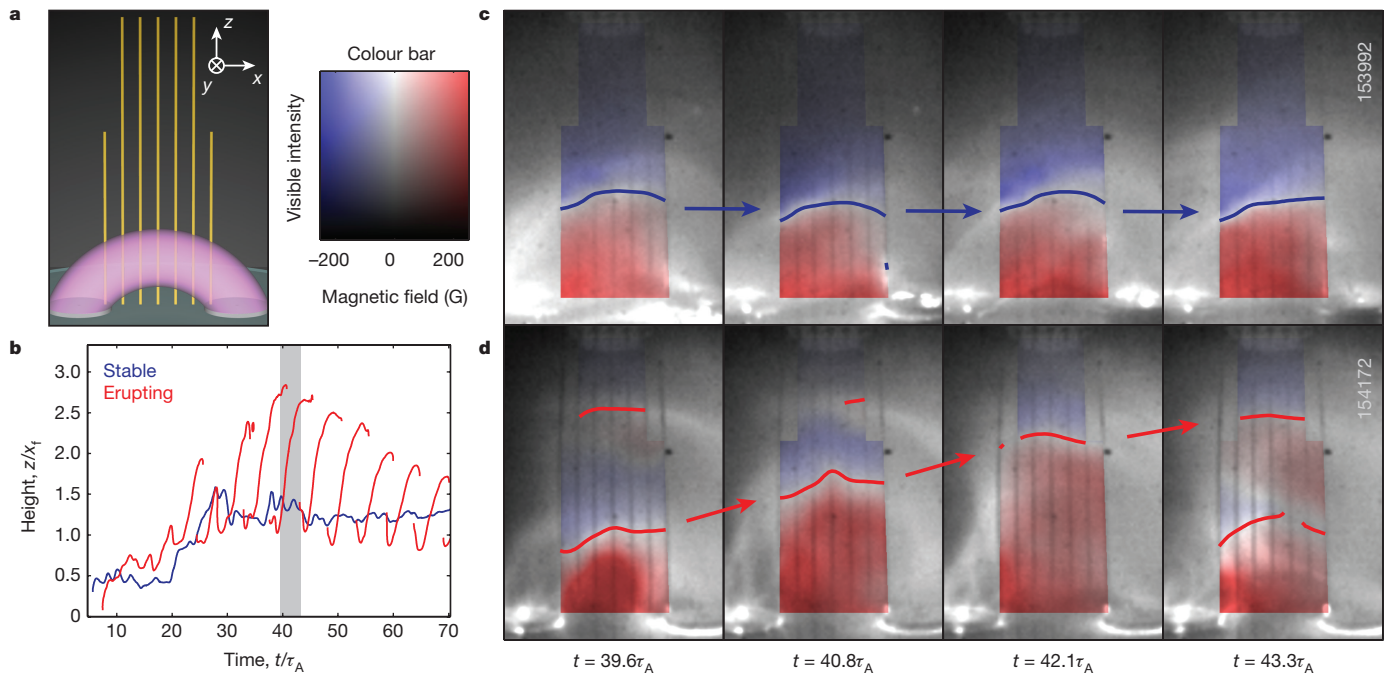


Figure 1 | Representative stable and erupting flux rope discharges. **a**, Experimental setup showing the arched flux rope (pink) attached to two conducting footpoints. The yellow vertical lines represent the *in situ* magnetic probes (see Methods). **b**, Height–time histories of the two flux rope discharges. The frame sequences in **c** and **d** are taken from the short time period shaded in grey. **c**, **d**, Frame sequences with the measured

out-of-plane magnetic field overlaid on corresponding fast camera visible light images (data ID numbers are shown on the right). The measured magnetic axis locations (the solid lines) are defined by the reversal of the out-of-plane magnetic field (see Methods). A video of the full discharge evolution is included as a Supplementary Video. τ_A , dynamic Alfvén time; x_f , footpoint separation distance; z , vertical height above the footpoints.

descriptions of the various current and field components). The internal profile of J_T rapidly transforms from nearly uniform to strikingly hollow during the failed torus event. This hollowing of the current profile is accompanied by a transient increase in the internal toroidal magnetic field, B_{Ti} (Fig. 3e). The toroidal field B_{Ti} and its associated poloidal currents, J_p , are self-generated by the plasma in order to achieve

a force-free state. Given that both the laboratory and solar flux ropes are magnetically rather than thermally dominated, the measured B_{Ti} is paramagnetic in nature (that is, it enhances rather than cancels the ambient guide field, B_g). As such, the poloidal currents, J_p , cross with the toroidal field, B_T , to produce a large, dynamic tension force that causes the eruption to fail (see Methods).

In the absence of substantial B_g , this tension force is much reduced. This leads to the eruptive behaviour shown in Extended Data Fig. 3, where the J_T profile remains relatively uniform throughout the event and the flux rope expands freely towards the wall of the machine. The observed rapid reformation of the flux rope after the eruption may differ from events in the solar corona. Assessing the impact of laboratory factors such as external inductance and boundary conditions on this phenomenon is an important topic for future work.

As a final step, we now quantitatively examine the magnetic forces acting on the flux rope. The three forces considered here are the hoop (F_h), strapping (F_s), and toroidal field tension (F_t) terms (see Methods and Extended Data Table 3). For the failed torus event in Fig. 3, all three force terms initially decline in magnitude (Fig. 3c). As the event proceeds, however, the tension force dramatically surges in magnitude, thereby halting the upward motion of the flux rope. For the eruptive event in Extended Data Fig. 3, on the other hand, all three force terms decline monotonically. The remarkable transient increase of the tension force in the failed event warrants further investigation. Figure 3b shows that there is a rapid conversion of poloidal to toroidal magnetic flux during the failed torus event. This flux conversion is the signature of a dynamic plasma relaxation event such as those observed in laboratory fusion devices²⁹.

Relaxation events occur because the plasma can find a lower energy state through internal reconfiguration rather than through external eruption. The traditional view is that the system ‘self-organizes’ to a lower energy state while conserving magnetic helicity, and that the underlying physical mechanism is magnetic reconnection³⁰. This reconnection is transient, three-dimensional, and internal to the flux rope, making it difficult to track experimentally. Nonetheless, the

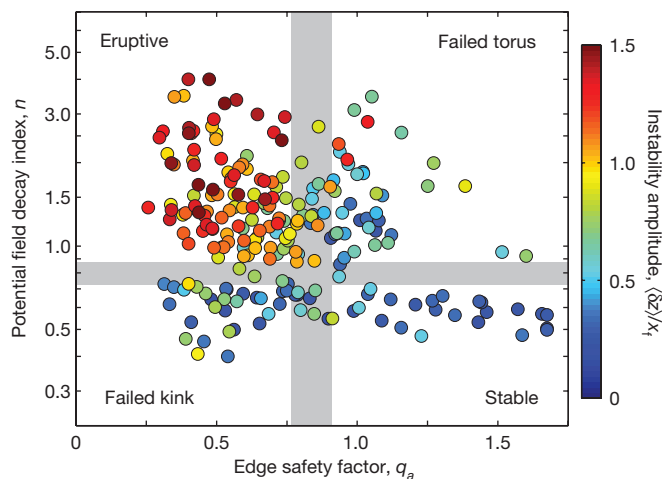


Figure 2 | The experimentally measured torus versus kink instability parameter space. The x axis represents the kink instability through the edge safety factor q_a (the inverse magnetic twist), while the y axis represents the torus instability through the potential field decay index n . Each data point is the mean of 2–5 flux rope plasma discharges with the same experimental parameters. A total of 806 flux rope plasma discharges are represented. The metric used here to quantify the eruptivity of each flux rope is the normalized spatial instability amplitude $\langle \delta z \rangle / x_f$ (see Methods). A value of $\langle \delta z \rangle / x_f < 0.5$ is stable, while $\langle \delta z \rangle / x_f > 1$ is clearly eruptive. The shaded boundaries, which are empirically identified, delineate the four distinct instability parameter regimes described in the text.

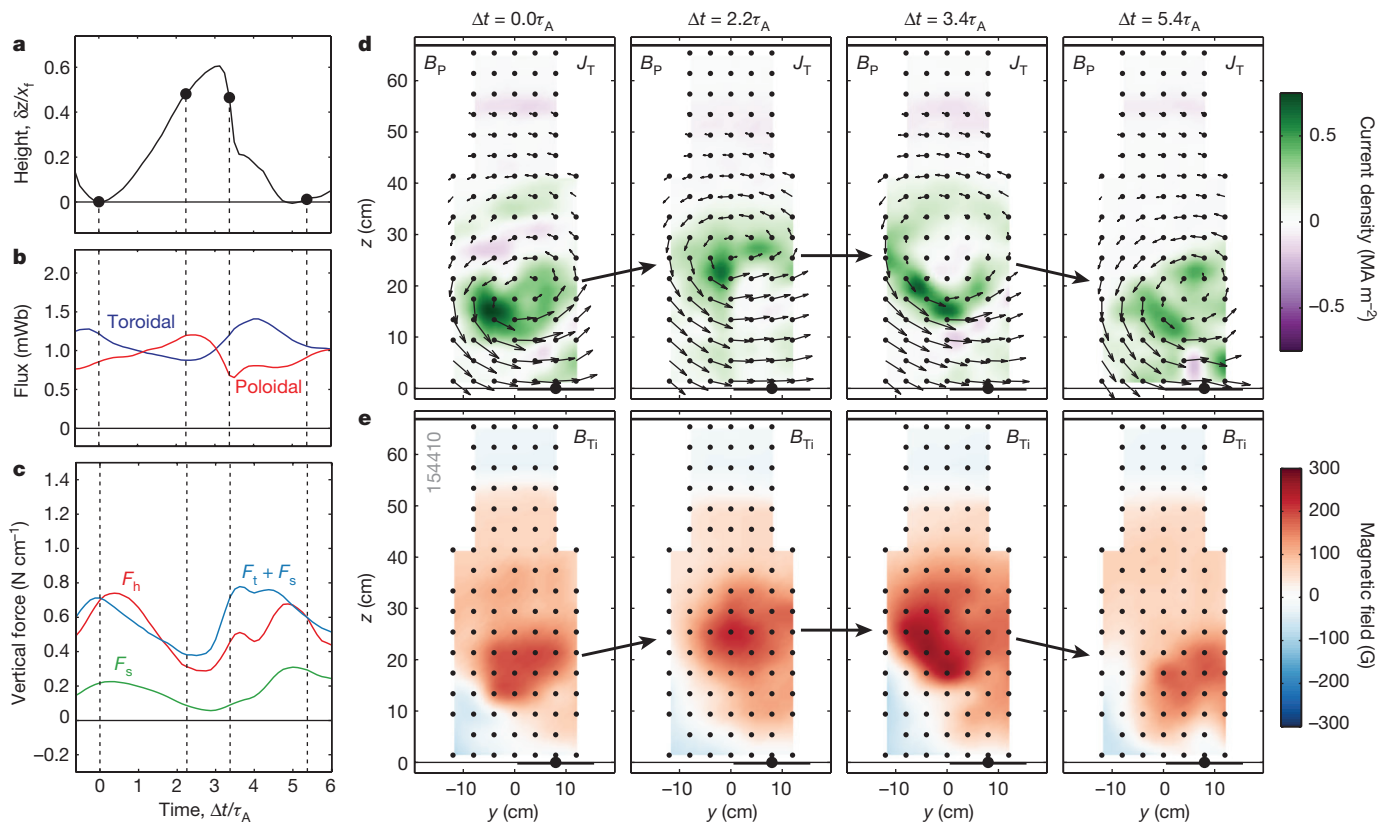


Figure 3 | Magnetic analysis of a characteristic failed torus event. See Extended Data Fig. 4b for the magnetic probe orientation. **a**, Relative perturbation amplitude showing that the flux rope initially expands before collapsing back downward. **b**, Time evolution of the toroidal and poloidal magnetic fluxes within the flux rope. **c**, Time evolution of the hoop (F_h),

strapping (F_s), and toroidal field tension (F_t) forces, showing the surge in the tension force that ultimately causes the event to fail. **d**, **e**, Sequenced spatial plots of the toroidal current density (J_T) and the internal toroidal field (B_{Ti}) showing the dramatic hollowing of J_T and the simultaneous transient increase in B_{Ti} (compare with Extended Data Fig. 3).

plasma's tendency to conserve helicity sheds light on the observed behaviour. Helicity characterizes the linkage between the poloidal and toroidal fluxes such that the product of the two is approximately conserved. Thus, in order to conserve helicity, the hollowing of the J_T profile, which reduces the poloidal flux in the rope, must be accompanied by a surge in the toroidal flux (and therefore a surge in the toroidal field tension force). Finally, we observe relaxation events only when the potential guide field is large enough to prevent the flux rope from kinking (that is, $q_a > 0.8$). When $q_a < 0.8$, on the other hand, self-organization fails because of the disruptive nature of the external kink mode.

With the laboratory results in hand, we now turn to their implications for eruptions in the solar corona. First, the existence of the failed torus regime implies that the onset of the torus instability is not a sufficient condition for eruption. Therefore, the toroidal field tension force that produces failed torus events must be added to the physical models that are used to study solar eruptions. Doing so presents a substantial challenge for two reasons.

First, because the toroidal field tension force dynamically surges during a failed torus event, time-resolved modelling of the flux rope is crucial. This rules out quasi-static nonlinear force-free field modelling, which has shown promise as a tool for understanding coronal configurations such as erupting sigmoids¹⁴. Second, the plasma relaxation events that enhance the toroidal field tension force are inherently three-dimensional²⁹. Therefore, the full line-tied geometry of the flux rope must be modelled in both time and space in order to resolve the physical mechanisms that define the failed torus regime. These difficult modelling requirements may explain why this regime has not been previously identified in numerical simulations.

Our results also have direct implications for remote observations of the corona. For example, the presence of a substantial guide magnetic

field in the potential field configuration of a given flux rope should indicate a reduced probability of eruption. This information can be obtained from relatively simple reconstructions of the flux rope's magnetic topology, even if a full model of the dynamically evolving magnetic field is not available. One promising candidate for study is the recent non-eruptive active region of the Sun's surface, NOAA AR 12192, which was one of the largest and longest-lived active regions of the space age. This region produced multiple large flares (it was the most prolific active region in solar cycle 24), but no coronal mass ejections were observed during its disk passage¹⁵. Preliminary inspection of the observational data shows that a number of the flares were associated with failed eruptions in the torus-unstable regime. If these events were indeed failed torus events, they may be explained by the toroidal field tension force mechanism identified here.

Finally, our results do not preclude the torus instability as an eruption mechanism for kink-stable flux ropes. Rather, they demonstrate that torus-driven eruptions can fail under certain conditions. Thus, comparing and contrasting the features of kink-stable flux ropes that do erupt with those that fail is a key next step towards a comprehensive understanding of the flux rope instability parameter space.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 May; accepted 29 October 2015.

- Kunow, H., Crooker, N. U., Linker, J. A., Schwenn, R. & von Steiger, R. (eds) *Coronal Mass Ejections* Ch. 2, 12 (Springer, 2006).
- Kuperus, M. & Raadu, M. A. The support of prominences formed in neutral sheets. *Astron. Astrophys.* **31**, 189–193 (1974).
- Chen, J. Effects of toroidal forces in current loops embedded in a background plasma. *Astrophys. J.* **338**, 453–470 (1989).

4. Titov, V. S. & Démoulin, P. Basic topology of twisted magnetic configurations in solar flares. *Astron. Astrophys.* **351**, 707–720 (1999).
5. Amari, T., Canou, A. & Aly, J.-J. Characterizing and predicting the magnetic environment leading to solar eruptions. *Nature* **514**, 465–469 (2014).
6. Ji, H., Wang, H., Schmahl, E. J., Moon, Y.-J. & Jiang, Y. Observations of the failed eruption of a filament. *Astrophys. J.* **595**, L135–L138 (2003).
7. Török, T. & Kliem, B. Confined and ejective eruptions of kink-unstable flux ropes. *Astrophys. J.* **630**, L97–L100 (2005).
8. Joshi, N. C. *et al.* Confined partial filament eruption and its reformation within a stable magnetic flux rope. *Astrophys. J.* **787**, 11 (2014).
9. Forbes, T. G. & Isenberg, P. A. A catastrophe mechanism for coronal mass ejections. *Astrophys. J.* **373**, 294–307 (1991).
10. Kliem, B. & Török, T. Torus instability. *Phys. Rev. Lett.* **96**, 255002 (2006).
11. Fan, Y. & Gibson, S. E. Onset of coronal mass ejections due to loss of confinement of coronal flux ropes. *Astrophys. J.* **668**, 1232–1245 (2007).
12. Liu, Y. Magnetic field overlying solar eruption regions and kink and torus instabilities. *Astrophys. J.* **679**, L151–L154 (2008).
13. Démoulin, P. & Aulanier, G. Criteria for flux rope eruption: non-equilibrium versus torus instability. *Astrophys. J.* **718**, 1388–1399 (2010).
14. Savcheva, A., Parlat, E., van Ballegoijen, A., Aulanier, G. & DeLuca, E. Sigmoidal active region on the sun: comparison of a magnetohydrodynamical simulation and a nonlinear force-free field model. *Astrophys. J.* **750**, 15 (2012).
15. Sun, X. *et al.* Why is the great solar active region 12192 flare-rich but CME-poor? *Astrophys. J.* **804**, L28 (2015).
16. Myers, C. E. *Laboratory Study of the Equilibrium and Eruption of Line-Tied Magnetic Flux Ropes in the Solar Corona*. <http://arks.princeton.edu/ark:/88435/dsp01dv13zw44b>, PhD thesis, Princeton Univ. (2015).
17. Hansen, J. F. & Bellan, P. M. Experimental demonstration of how strapping fields can inhibit solar prominence eruptions. *Astrophys. J.* **563**, L183–L186 (2001).
18. Soltwisch, H. *et al.* Flarelab: early results. *Plasma Phys. Contr. Fusion* **52**, 124030 (2010).
19. Tripathi, S. K. P. & Gekelman, W. Laboratory simulation of arched magnetic flux rope eruptions in the solar atmosphere. *Phys. Rev. Lett.* **105**, 075005 (2010).
20. Gold, T. & Hoyle, F. On the origin of solar flares. *Mon. Not. R. Astron. Soc.* **120**, 89–105 (1960).
21. Sakurai, T. Magnetohydrodynamic interpretation of the motion of prominences. *Publ. Astron. Soc. Jpn.* **28**, 177–198 (1976).
22. Hood, A. W. & Priest, E. R. Critical conditions for magnetic instabilities in force-free coronal loops. *Geophys. Astrophys. Fluid Dyn.* **17**, 297–318 (1981).
23. Mikić, Z., Schnack, D. D. & van Hoven, G. Dynamical evolution of twisted magnetic flux tubes. I—Equilibrium and linear stability. *Astrophys. J.* **361**, 690–700 (1990).
24. Török, T., Kliem, B. & Titov, V. S. Ideal kink instability of a magnetic loop equilibrium. *Astron. Astrophys.* **413**, L27–L30 (2004).
25. Kruskal, M. & Schwarzschild, M. Some instabilities of a completely ionized plasma. *Proc. R. Soc. Lond. A* **223**, 348–360 (1954).
26. Shafranov, V. The stability of a cylindrical gaseous conductor in a magnetic field. *Sov. J. At. Energy* **1**, 709–713 (1956).
27. Ryutov, D. D., Furno, I., Intrator, T. P., Abbate, S. & Madziwa-Nussinov, T. Phenomenological theory of the kink instability in a slender plasma column. *Phys. Plasmas* **13**, 032105 (2006).
28. Olmedo, O. & Zhang, J. Partial torus instability. *Astrophys. J.* **718**, 433–440 (2010).
29. Ji, H., Prager, S. C. & Sarff, J. S. Conservation of magnetic helicity during plasma relaxation. *Phys. Rev. Lett.* **74**, 2945–2948 (1995).
30. Taylor, J. B. Relaxation and magnetic reconnection in plasmas. *Rev. Mod. Phys.* **58**, 741–763 (1986).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. Cutler for constructing the flux rope experiment and for myriad technical contributions. We also thank F. Scotti and P. Sloboda for additional technical contributions and R. M. Kulsrud for theoretical discussions. This research is supported by Department of Energy (DoE) contract number DE-AC02-09CH11466 and by the National Science Foundation/DoE Center for Magnetic Self-Organization (CMSO).

Author Contributions C.E.M., M.Y. and H.J. designed the laboratory experiments. C.E.M., J.Y. and J.J.-A. carried out the experiments and processed the data. C.E.M., M.Y., H.J., J.Y., W.F. and J.J.-A. interpreted the laboratory results. A.S. and E.E.DeL. placed the laboratory results in the context of solar observations and modelling. C.E.M. analysed the laboratory data, prepared the figures, and wrote the manuscript. All authors contributed to the revision of the manuscript.

Author Information The digital data for this paper can be found at <http://arks.princeton.edu/ark:/88435/dsp01j3860933c>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.E.M. (cmeyers@pppl.gov).

METHODS

Candidate solar eruption mechanisms. Ideal magnetohydrodynamic instabilities such as the torus and kink instabilities are central to the standard storage-and-release model of solar flares and coronal mass ejections¹. In addition to such ideal instabilities, the non-ideal process of magnetic reconnection is routinely invoked to explain various observed solar flare and coronal mass ejection features. For example, reconnection produces flare emission beneath the expanding/rising flux rope and contributes to the evolution of the flux rope height³¹. Reconnection is also the central driving mechanism in some coronal mass ejection initiation models³². Magnetohydrodynamic simulations and data-driven modelling have shown, however, that the torus instability plays a crucial part in driving magnetic flux ropes to erupt, even in the presence of magnetic reconnection¹⁴. Accordingly, our flux rope experiments are designed to identify the stability boundaries for the triggering of candidate ideal instability eruption mechanisms.

The torus instability is triggered by an imbalance in the vertical forces acting on the flux rope plasma¹⁰. The traditional forces considered for the torus instability are (1) the upward ‘hoop’ force F_h , which is the Lorentz force between the toroidal (axial) flux rope current and its own poloidal (azimuthal) magnetic field; and (2) the downward ‘strapping’ force F_s , which is the Lorentz force between the same toroidal current and the potential strapping field (see the Methods subsection ‘Magnetic force analysis’). Analysis of Shafranov’s toroidal equilibrium equations³³ reveals that the torus instability threshold can be expressed analytically in terms of the potential field ‘decay index’^{10,34}:

$$n(z) = -\frac{z}{|B_{\text{pot}}|} \frac{\partial |B_{\text{pot}}|}{\partial z} \quad (1)$$

where B_{pot} is the potential magnetic field and z is the height above the solar surface. A larger value of n indicates a more quickly decaying potential field. For a toroidally symmetric, large-aspect-ratio flux rope, the torus instability criterion^{10,34} reduces to $n \geq 3/2$, which is a remarkably concise result given the complexity of the system. Much effort has been expended to more accurately determine the torus threshold for the realistic line-tied conditions of the solar corona, but a wide range of estimates remain^{13,16,28,35}.

The kink instability^{20–24}, on the other hand, is triggered when the magnetic twist at the edge of the flux rope (that is, the poloidal angle through which an edge magnetic field line rotates as it transits the toroidal length of the flux rope) exceeds a critical threshold^{25,26}. The analytical kink onset condition is often given in terms of the edge safety factor^{25–27}, q_a , which is defined as the inverse of the edge magnetic twist, ι_a :

$$q_a \equiv \frac{2\pi}{\iota_a} = \frac{d\Phi_T}{d\psi_p} \Big|_{r=a} \approx \frac{2\pi a}{L} \frac{B_{Ta}}{B_{Pa}} \quad (2)$$

Here, Φ_T is the enclosed toroidal magnetic flux, ψ_p is the enclosed poloidal magnetic flux, r is the minor radial coordinate, and a is the minor radius of the flux rope. In the latter expression, L is the rope length, B_{Ta} is the edge toroidal field strength, and B_{Pa} is the edge poloidal field strength. The well known Kruskal–Shafranov kink criterion^{25–27} predicts instability for $q_a \leq 1$, but numerical analyses of arched, line-tied flux ropes at finite aspect ratio^{22,24} have predicted a more stable criterion of $q_a \leq 0.8$. Previous laboratory experiments on linear^{36–38} and arched³⁹ line-tied flux ropes have demonstrated the importance of the line-tied boundary conditions to the kink stability criterion. In spite of these efforts, the combined stability against both torus and kink perturbations in the two-dimensional n versus q_a parameter space has not been well explored.

Experimental setup and solar relevance. Our experiments are conducted in the Magnetic Reconnection Experiment (MRX)⁴⁰ at Princeton Plasma Physics Laboratory. To produce solar-relevant line-tied magnetic flux ropes, the MRX device is substantially modified from its standard operating mode¹⁶. In particular, its magnetic-reconnection-producing ‘flux cores’ are removed and replaced with a custom-built flux rope apparatus that contains the following: (1) two electrodes that serve as the flux rope footpoints; (2) two sets of magnetic field coils inside the vessel that produce the guide and strapping potential magnetic field; and (3) a glass substrate that physically separates the $z > 0$ plasma region from the $z < 0$ field coil region (see Extended Data Fig. 1). The two electrodes are circular copper discs with a footpoint radius of $a_f = 7.5$ cm and a horizontal separation distance of $2x_f = 36$ cm. The entire flux rope apparatus is housed within a cylindrical stainless steel vacuum vessel that is evacuated to $p \approx 10^{-6}$ Torr. Finally, two additional sets of magnetic field coils located outside the vessel are used to adjust the guide and strapping field spatial profiles.

Before a flux rope plasma can be produced in the experiment, the desired potential magnetic field configuration must be created. This is accomplished by energizing the four independent magnetic field coil sets introduced above.

Each potential field component (guide or strapping) is produced by superposing the fields from two of the four available coil sets (one inside the vessel and one outside the vessel per field component). This superposition provides two degrees of freedom for each field component that are typically used to independently set the field strength and the field decay index (see equation (1)). The independent control of these two parameters for both the guide and strapping fields facilitates a systematic exploration of the torus versus kink instability parameter space.

Once a given potential field configuration has been selected, a precisely timed sequence of events is initiated. First, the potential magnetic field coils are energized to their requested settings and held there for the duration of the discharge. In practice, the potential field ramp is completed 7 ms before the formation of the flux rope plasma. This is more than twice the inductive skin time of the vessel wall and of the copper electrodes ($\tau_w \approx \tau_f \approx 3$ ms), such that any induced eddy currents decay away before the plasma is formed. Next, neutral gas, typically hydrogen, is injected into the vessel to provide a particle source for the plasma. The gas is injected at both the vessel wall and directly at the cathode surface to ensure consistent plasma breakdown at reasonable fill pressures and firing voltages ($p \approx 10$ mTorr; $V \approx 4$ kV). Finally, a charged capacitor bank is connected across the electrodes to break down the neutral gas into an arc discharge plasma. As electric current and therefore free magnetic energy is slowly injected into the system, the pre-existing potential magnetic field lines are twisted into a magnetic flux rope. This procedure is repeated thousands of times over the course of the experimental campaign to generate flux ropes with a wide range of equilibrium and stability properties.

The typical parameters of our laboratory flux ropes are displayed in Extended Data Table 1. These laboratory parameters can be used to compute key dimensionless physics parameters that justify the relevance of our laboratory experiments to storage-and-release eruptions in the solar corona (see Extended Data Table 2). First, a strict timescale ordering must be satisfied. In particular, the abovementioned driving timescale, τ_D , must be both substantially longer than the dynamic Alfvén timescale, τ_A , and substantially shorter than the resistive timescale, τ_R . The separation between τ_A and τ_D satisfies the storage-and-release requirement, while the separation between τ_D and τ_R respects the high conductivity of the solar corona.

Additionally, for the physical phenomena observed in the laboratory to be independent of scale (and therefore be applicable to the corona), the laboratory plasma must reside in the magnetohydrodynamic regime. Such extrapolation is possible because magnetohydrodynamics has no fundamental spatial length scale⁴¹. The magnetohydrodynamic nature of a given plasma is characterized by the remaining parameters in Extended Data Table 2. First, $\rho_i/a \ll 1$ indicates that the ratio of the Larmor radius of individual ions to the flux rope minor radius is small, such that scale-dependent finite Larmor radius effects are negligible. Second, $\lambda_{ei}/L \ll 1$ indicates that the plasma collisionality is high, such that the fluid approximation employed by magnetohydrodynamic is valid. Third, the Lundquist number $S \gg 1$ is large, such that magnetic field lines are frozen into the plasma and ideal magnetohydrodynamic instabilities such as the kink and torus instabilities will govern the behaviour of the system. Fourth, the ionization fraction, $n_e/(n_e + n_i)$, indicates that the laboratory plasma is ionized sufficiently for magnetohydrodynamic rather than neutral physics to dominate. Finally, the plasma $\beta \ll 1$ indicates that the plasma is magnetically rather than thermally dominated. This combination of dimensionless parameters justifies the application of our laboratory experiments to the solar eruption problem.

Laboratory diagnostics. Two primary diagnostics are used in our experiments: fast visible-light cameras and *in situ* magnetic probes. Data from both diagnostics are compared in Fig. 1. The fast cameras are used to qualitatively assess the location and performance of the arc discharge plasmas. They are Vision Research Phantom v710 monochrome cameras operated with a 1- μ s exposure at 270,000 frames per second (~ 3 - μ s, 1- τ_A cadence). The collected light spans the visible spectrum, with the primary contribution coming from the H α hydrogen neutral line. The dominance of neutral light in these images makes them fundamentally different from the extreme-ultraviolet images of the solar corona that are acquired by instruments such as the Atmospheric Imaging Assembly (AIA) aboard the Solar Dynamics Observatory (SDO)⁴².

The *in situ* magnetic probes, on the other hand, directly measure the internal magnetic structure of the flux rope plasma. Each probe is constructed from a long, thin glass tube (64 cm long, 0.7 cm in diameter) that houses up to 51 miniature magnetic pickup coils that are distributed along its length. These pickup coils each measure the time derivative of one component of the vector magnetic field, and the resulting signals are integrated to measure the magnetic field as a function of time. The pickup coils are grouped in orthogonal triplets to measure the complete vector field at each spatially distributed location. Seven such probes housing approximately 300 total pickup coils are inserted into the plasma in order to map out the magnetic field at more than 100 locations in a two-dimensional plane. The triplets within each probe are separated vertically at 4 cm intervals, and the seven probes

are separated horizontally by 4 cm to produce a $4\text{ cm} \times 4\text{ cm}$ measurement grid over a $24\text{ cm} \times 64\text{ cm}$ cross-section of the plasma. As shown in Extended Data Fig. 4, this two-dimensional plane can be oriented parallel to or orthogonal to the flux rope axis. Sample magnetic field measurements for each case are also shown, with the colour representing the out-of-plane field and the vectors representing the in-plane field. Both the arched shape of the flux rope and its quasi-circular cross-section are clearly visible in these data. The magnetic field data are digitized at 2.5 MHz ($0.4\text{ }\mu\text{s}$, $0.1\text{ }\tau_A$ timebase). As such, the instabilities studied here are well resolved temporally. Though the magnetic probes are inserted directly into the plasma, they are thin and non-conducting and are therefore largely non-perturbative. Their use in MRX for detailed physics studies is well established⁴³.

Height–time evolution and instability parameter space analysis. To characterize the behaviour of a given flux rope plasma, the spatially distributed magnetic field data acquired during the discharge can be reduced to a ‘height–time’ plot that succinctly tracks the evolution of the flux rope magnetic axis. This is accomplished by selecting a single vertical magnetic probe from the array and extracting the measured $B_y(t, z)$ data. The B_y field component is the superposition of the ‘internal’ poloidal field produced by the plasma, B_{pi} , and the external strapping field, B_s . Its reversal point at $B_y(z, t) = 0$ therefore constitutes a measurement of the magnetic axis of the flux rope. Four sample height–time plots, one from each of the four instability regimes identified in Fig. 2, are shown in Extended Data Fig. 5. The colour in each height–time plot represents $B_y(z, t)$, with the black line indicating the measured magnetic axis location. The qualitative differences between the different instability regimes are clearly visible in these plots. To arrive at the more quantitative assessment of the instability parameter space presented in Fig. 2, however, the height–time data must be further reduced.

In our experiments, we use three scalar quantities to summarize the performance of a given flux rope plasma: (1) the edge safety factor, q_{a1} ; (2) the field decay index, n ; and (3) the spatial instability amplitude, $\langle \delta z \rangle / x_f$. The first two parameters place the plasma within the torus versus kink instability parameter space, while the third is a metric developed to quantify the eruptivity of a given flux rope. In each discharge, q_a and n are evaluated at the maximum of the $\langle z_{\text{apex}} \rangle$ waveform, which tracks the time-averaged height of the flux rope apex (see Extended Data Fig. 5). The evaluation of n via equation (1) is straightforward given that the potential field magnitude, $|B_{\text{pot}}|$, is well defined by the geometry of the magnetic field coils in the experiment.

To evaluate $q_a \approx 2\pi a B_{T0} / L B_{p0}$ using equation (2), on the other hand, the footpoint values of the minor radius and the magnetic fields are used: $a = a_f$, $B_{T0} = B_{gf}$, and $B_{p0} = B_{pf} \approx \mu_0 I_T / 2\pi a_f$, where I_T is the toroidal flux rope current. The length of the rope, L , is approximated here using a ‘shifted-circle’ model for the rope axis^{3,16} that depends only on the apex height, $\langle z_{\text{apex}} \rangle$, and the footpoint separation distance, x_f . This approximation for q_a assumes that toroidal flux is conserved along the length of the flux rope. It can have errors of up to 10%, however, which are mostly caused by uncertainty in the fraction of the measured capacitor bank current that is carried in the flux rope. Based on magnetic probe measurements, this fraction is typically 90%. The final step is to evaluate the instability amplitude metric, $\langle \delta z \rangle / x_f$. Here, the dynamic spatial amplitude $\langle \delta z \rangle$ is defined as the maximum of the envelope of the dynamic motion of the magnetic axis. The relevant values of q_a , n , and $\langle \delta z \rangle / x_f$ are listed in Extended Data Fig. 5c. These values show that the instability amplitude provides a quantitative assessment of the qualitatively disparate behaviours of the four flux rope discharges in Extended Data Fig. 5b. Finally, in order to produce the parameter space scatterplot in Fig. 2, the data from multiple flux rope plasmas with the same experimental parameters are combined. Each data point in Fig. 2 contains the mean of 2–5 flux rope plasma discharges such that more than 800 discharges are represented.

Magnetic force analysis. The magnetic probe data are also used to directly measure the magnetic forces acting on the line-tied flux rope. These force measurements are used to demonstrate the key role of the toroidal field tension force in the failed torus regime. The forces in a low- β plasma (one with negligible thermal pressure) are dominated by magnetic $\mathbf{J} \times \mathbf{B}$ Lorentz forces, where \mathbf{J} is the current density and \mathbf{B} is the magnetic field. Here, the total force density $\mathbf{J} \times \mathbf{B}$ is decomposed into three key contributions: (1) the hoop force, F_h ; (2) the strapping force, F_s ; and (3) the tension force, F_t (see Extended Data Table 3). The hoop force pushes the flux rope plasma upward, while the strapping and tension forces push downward and work together to confine the rope.

The first step in evaluating the three force terms described above is to decompose the magnetic field and current density into the individual components that contribute to each force term (see Extended Data Table 3). Sample magnetic field and current density measurements are shown in Extended Data Fig. 6. The computation of J_θ from the B_{T1} data requires a measurement of the toroidal curvature of the rope (see below). The final output of the field and force decomposition in Extended Data Table 3 is the set of force densities f_h , f_s , and f_t . These quantities are ‘force densities’ rather than forces because they have units of force per volume.

The forces plotted in Fig. 3 and Extended Data Fig. 3, on the other hand, are the forces per unit length, F_h , F_s , and F_t , that are integrated from the abovementioned force density terms. It is important to note that the tension force density, f_t , actually contains both magnetic tension and pressure contributions. The tension contribution is derived from the toroidal curvature of the magnetic field in the arched flux rope, and at large aspect ratio its leading term is proportional to $B_{T1} B_{T1} / R$, where R is the radius of curvature of the flux rope¹⁶. The pressure contribution, on the other hand, is derived from gradients in the internal toroidal field, B_{T1} . In practice, the tension contribution to f_t dominates the pressure contribution in the failed torus regime. As such, here we refer to f_t as simply the toroidal field tension force to avoid unnecessarily complicating the physics discussion.

As noted above, the force densities must be integrated over the cross-section of the flux rope. This converts the force densities, f_i , to the forces per unit length, F_i , that are plotted in Fig. 3c and Extended Data Fig. 3c. The cross-section integral takes the form:

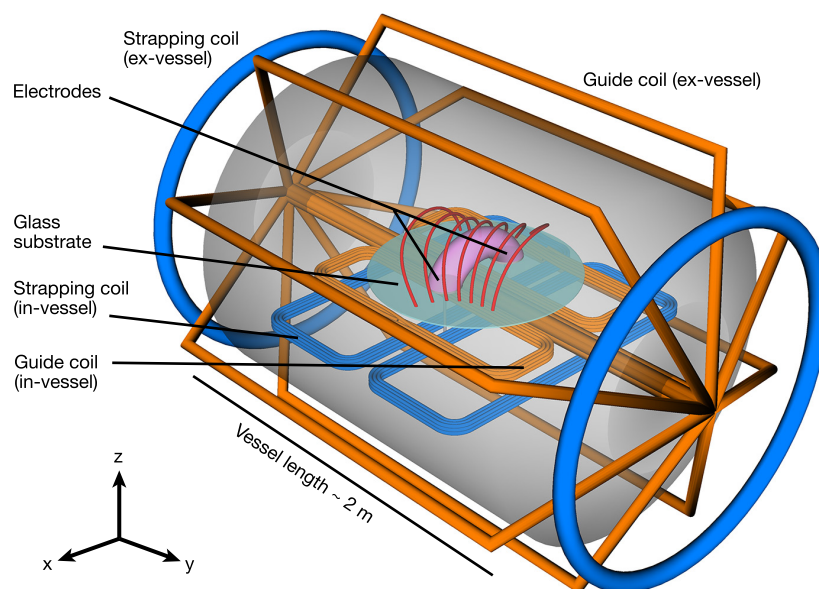
$$F(z_{\text{apex}}) = \frac{1}{R_{\text{apex}}} \int_0^{2\pi} d\theta \int_0^a dr [r h_T(z) f(r, \theta)] \quad (3)$$

where R_{apex} is the radius of curvature at the flux rope apex, (r, θ) are cylindrical coordinates in the (y, z) plane, $a(\theta)$ is the flux rope minor radius, and h_T is the toroidal curvilinear scale factor that accounts for the toroidal curvature of the flux rope. The curvilinear scale factor is directly measured from flux rope plasmas with the probe array aligned in the toroidal cross-section (see Extended Data Fig. 4). The resulting curvature measurements are then used to analyse the magnetic forces in equivalent flux rope plasmas with the probe array aligned in the poloidal cross-section¹⁶. The remaining quantity in equation (3) is the minor radius $a(\theta)$, which sets the extent of the flux rope cross-section. This quantity is obtained via the poloidal flux function of the flux rope $\psi(y, z)$. The flux function is obtained by line-integrating the measured poloidal magnetic field components as follows:

$$\psi(y, z) = - \int_{C_y} dy h_T B_z + \int_{C_z} dz h_T B_y \quad (4)$$

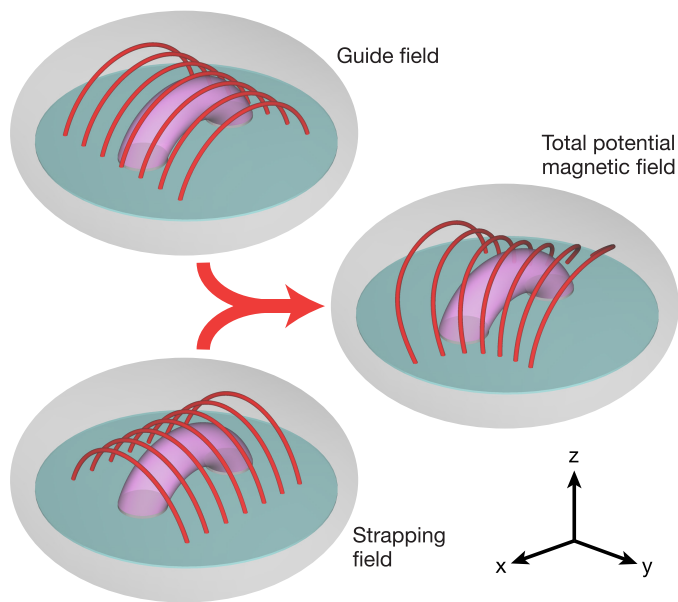
where B_y and B_z are the in-plane components of the poloidal field and C_y and C_z are the paths of integration along each direction. By construction, the integration is path independent. Contours of the resulting poloidal flux function are shown in blue on the left-hand side of Extended Data Fig. 6. The minor radius $a(\theta)$, shown in red, is defined by the flux function contour that encloses $\sim 90\%$ of the total current that is fed to the electrodes. With the minor radius now defined, the three forces per unit length can be computed at each instant in time. These integration techniques are also used to evaluate the toroidal and poloidal magnetic fluxes that are plotted in Fig. 3 and Extended Data Fig. 3. An extensive analysis of the equilibrium force balance in non-erupting flux ropes benchmarks the strapping force measured with these techniques to within 5% of analytical expectations. Furthermore, a force-free equilibrium is measured to within $\pm 15\%$ of the hoop force magnitude over an ensemble of hundreds of non-erupting flux ropes¹⁶. These results give confidence in the force measurements presented in Fig. 3 and Extended Data Fig. 3.

31. Moore, R. L., Sterling, A. C., Hudson, H. S. & Lemen, J. R. Onset of the magnetic explosion in solar flares and coronal mass ejections. *Astrophys. J.* **552**, 833–848 (2001).
32. Antiochos, S. K., DeVore, C. R. & Klimchuk, J. A. A model for solar coronal mass ejections. *Astrophys. J.* **510**, 485–493 (1999).
33. Shafranov, V. Plasma equilibrium in a magnetic field. *Rev. Plasma Phys.* (ed. Leontovich, M. A.) **2**, 103–152 (1966).
34. Bateman, G. *MHD Instabilities* (MIT Press, 1978).
35. Kliem, B., Lin, J., Forbes, T. G., Priest, E. R. & Török, T. Catastrophe versus instability for the eruption of a toroidal solar magnetic flux rope. *Astrophys. J.* **789**, 46 (2014).
36. Hsu, S. C. & Bellan, P. M. Experimental identification of the kink instability as a poloidal flux amplification mechanism for coaxial gun spheromak formation. *Phys. Rev. Lett.* **90**, 215002 (2003).
37. Furno, I. et al. Current-driven rotating-kink mode in a plasma column with a non-line-tied free end. *Phys. Rev. Lett.* **97**, 015002 (2006).
38. Bergerson, W. F. et al. Onset and saturation of the kink instability in a current-carrying line-tied plasma. *Phys. Rev. Lett.* **96**, 015004 (2006).
39. Oz, E. et al. Experimental verification of the Kruskal-Shafranov stability limit in line-tied partial-toroidal plasmas. *Phys. Plasmas* **18**, 102107 (2011).
40. Yamada, M. et al. Study of driven magnetic reconnection in a laboratory plasma. *Phys. Plasmas* **4**, 1936–1944 (1997).
41. Bellan, P. M. & Hansen, J. F. Laboratory simulations of solar prominence eruptions. *Phys. Plasmas* **5**, 1991–2000 (1998).
42. Lemen, J. R. et al. The Atmospheric Imaging Assembly (AIA) on the Solar Dynamics Observatory (SDO). *Sol. Phys.* **275**, 17–40 (2012).
43. Yoo, J., Yamada, M., Ji, H. & Myers, C. E. Observation of ion acceleration and heating during collisionless magnetic reconnection in a laboratory plasma. *Phys. Rev. Lett.* **110**, 215007 (2013).

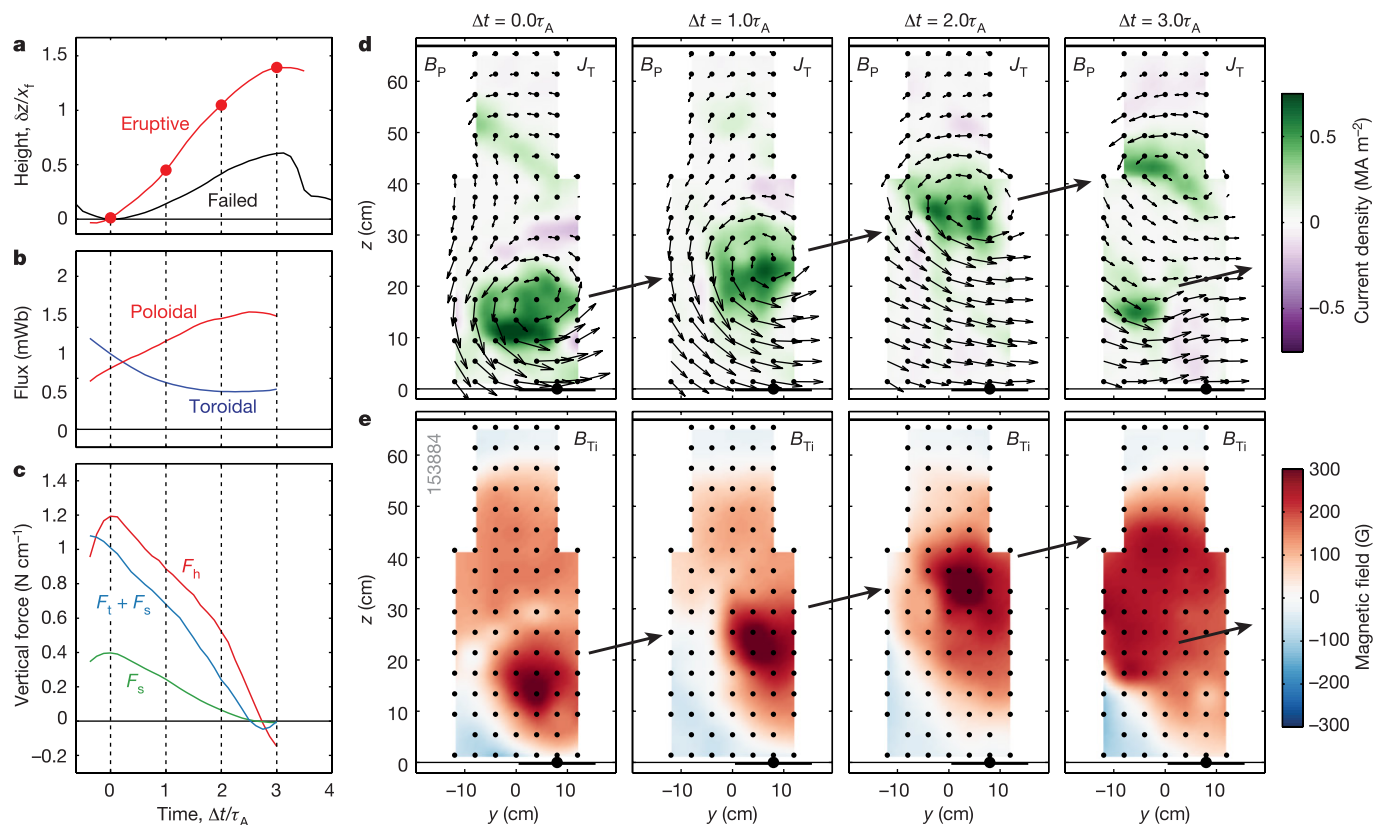


Extended Data Figure 1 | Experimental setup. A plasma arc (pink) is maintained between two electrodes that are mounted on a glass substrate. The electrodes, which serve as the flux rope footpoints, are horizontally separated by $2x_f = 36$ cm, and they have a minor radius of $a_f = 7.5$ cm. The vertical distance from these footpoints to the vessel wall is $z_w \approx 70$ cm.

Four magnetic field coil sets (two inside the vessel, two outside) work in concert to produce a variety of potential magnetic field configurations. More specifically, the two orange coil sets are used to produce the guide potential field, while the two blue coil sets are used to produce the strapping potential field.



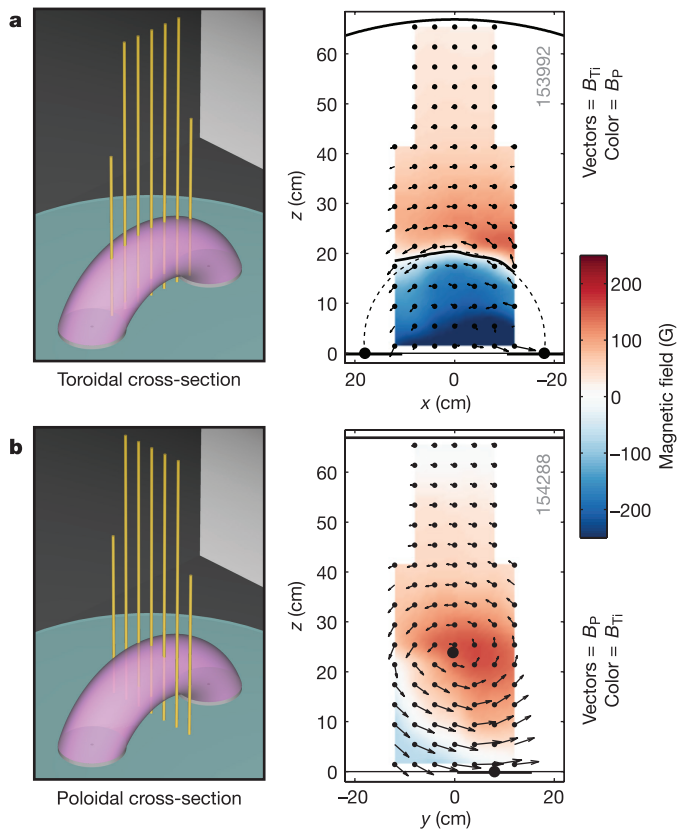
Extended Data Figure 2 | Components of the potential magnetic field configuration. The strapping field runs perpendicular to the flux rope axis and produces the well known strapping force, whose rapid spatial decay can trigger the torus instability. The guide field, on the other hand, runs toroidally along the flux rope axis. It stabilizes the kink instability and generates a confining magnetic tension force. The total potential magnetic field, which is the superposition of the guide and strapping field contributions, is obliquely aligned to the flux rope.



Extended Data Figure 3 | Magnetic field analysis of a characteristic eruptive event. **a**, The spatial evolution of the eruptive perturbation (red), with the failed torus event from Fig. 3a for comparison (black). **b**, Evolution of the poloidal and toroidal magnetic fluxes. Note the monotonic evolution of both fluxes. **c**, Hoop (F_h), strapping (F_s),

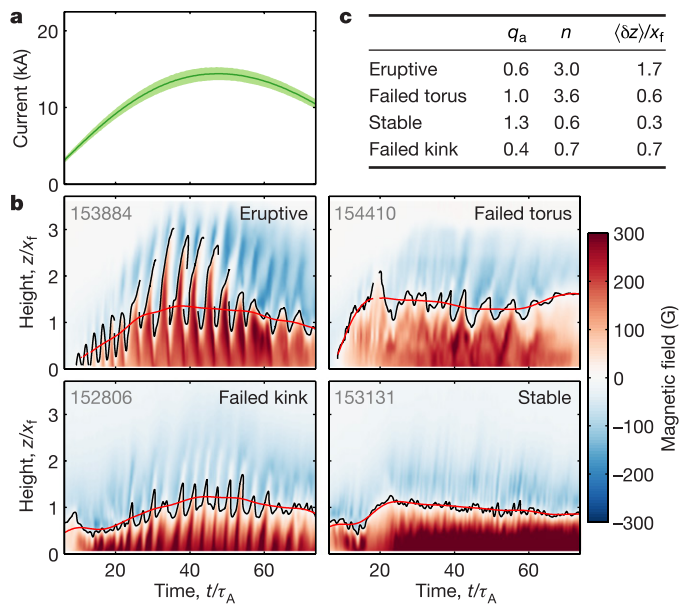
and tension (F_t) force evolution, which are also strictly monotonic.

d, e, Sequent J_T and B_{Ti} evolution. Note that the current profile remains uniform and rises steadily towards the wall of the machine. A new flux rope is forming at low altitude in the final frame.

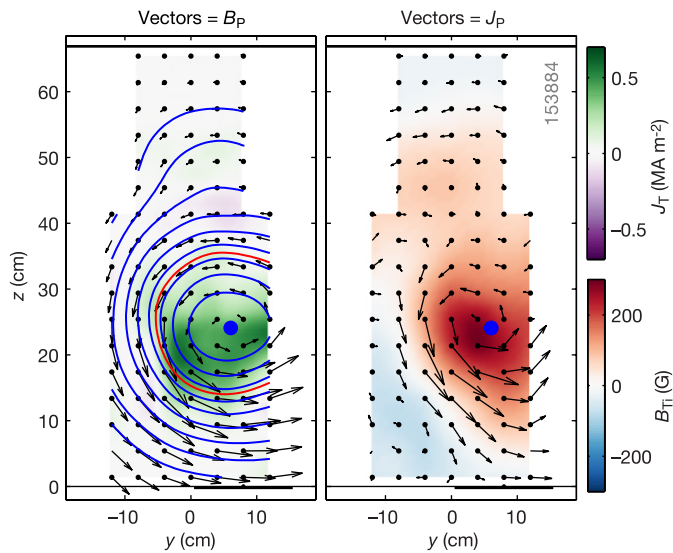


Extended Data Figure 4 | Sample *in situ* magnetic field measurements.

Seven linear magnetic field probes (yellow) are inserted vertically into the flux rope plasma. The alignment of the two-dimensional probe plane is either (a) parallel to the footpoint axis or (b) perpendicular to it. In the sample data, the colour represents the out-of-plane field, while the vectors represent the in-plane field. The position of the magnetic axis in the toroidal cross-section (the solid black line) is determined by the reversal in the out-of-plane poloidal magnetic field, B_y . The position of the magnetic axis in the poloidal cross-section is defined as the O-point in the circulating in-plane field (B_y, B_z). The out-of-plane field in the latter case is the ‘internal’ toroidal field of the flux rope B_{Ti} , which is paramagnetic in nature.



Extended Data Figure 5 | Height-time plots from four representative flux rope discharges. **a**, Mean toroidal plasma current waveform showing that the plasma current is nearly the same in all four cases (the light green band is the standard deviation). **b**, Four sample height-time plots, one from each of the four stability regimes identified in Fig. 2. The magnetic axis position (the black line) is defined by the zero-crossing in the $B_y(t, z)$ data, which is shown in colour. The red line in each frame is the time-averaged height of the flux rope apex $\langle z_{\text{apex}} \rangle$. This waveform provides the height at which q_a and n are measured in each discharge. **c**, Table of extracted flux rope parameters for each discharge.



Extended Data Figure 6 | Magnetic field and current density data for computing flux rope forces. The probe array is aligned as shown in Extended Data Fig. 4b. In the left panel, the colour is the toroidal current density, J_T , and the vectors are the poloidal magnetic field, B_p . In the right panel, the colour is the internal toroidal field B_{Ti} , and the vectors are the poloidal current density J_p . With all components of \mathbf{J} and \mathbf{B} measured, the force densities listed in Extended Data Table 3 can be readily computed. The contours in the left panel are contours of the poloidal flux function $\psi(y, z)$ (see equation (4)). The minor radius of the rope $a(\theta)$ is defined by the poloidal flux contour shown in red (see Methods).

Extended Data Table 1 | Laboratory flux rope parameters

Laboratory parameter	Symbol	Value	Units
Magnetic field strength	B	300–500	G
Neutral density	n_n	$\sim 5 \times 10^{14}$	cm^{-3}
Electron density (approx.)	n_e	$5 \times 10^{13} - 1 \times 10^{14}$	cm^{-3}
Electron temperature (approx.)	T_e	3–5	eV
Footpoint-to-footpoint scale length	L	0.5	m
Alfvén velocity	v_A	65–150	km/s
Alfvén transit time	τ_A	3–8	μs
Footpoint driving time	τ_D	~ 150	μs
Resistive diffusion time (Spitzer)	τ_R	0.8–2	ms

The quoted magnetic field strength, B , represents the footpoint-to-footpoint average along the rope. The electron density, n_e , and temperature, T_e , are approximate, owing to the limited availability of Langmuir probe data from these arc discharge plasmas. The characteristic footpoint driving time, τ_D , is set by the capacitance, inductance and resistance of the combined capacitor bank and plasma arc circuit. The laboratory parameters in this table are used to compute the related dimensionless parameters in Extended Data Table 2.

Extended Data Table 2 | Comparison of solar and laboratory dimensionless parameters

Dimensionless parameter	Symbol	Solar	Laboratory
Footpoint driving time / Alfvén transit time	τ_D/τ_A	100–10 ⁴	20–50
Footpoint driving time / resistive diffusion time	τ_D/τ_R	10 ^{−7}	~0.1
Ion gyroradius / minor radius	ρ_i/a	10 ^{−6}	0.05
Electron mean free path / plasma length	λ_{ei}/L	10 ^{−2}	10 ^{−3} –10 ^{−2}
Lundquist number	S	10 ⁴ –10 ¹²	100–500
Ionisation fraction	$n_e/(n_e+n_n)$	50–100%	10–20%
Plasma beta (thermal pressure / magnetic pressure)	β	~1%	2–20%

While the laboratory experiments are not able to replicate the extreme parameters of the corona, they do satisfy the key dimensionless limits required to produce storage-and-release eruptions that are driven by ideal magnetohydrodynamic instabilities (see Methods).

Extended Data Table 3 | Decomposition of magnetic field, current density, and force terms

Quantity	Symbol	Expression
Strapping magnetic field (potential)	\mathbf{B}_s	—
Internal poloidal magnetic field (flux rope)	\mathbf{B}_{Pi}	—
Guide magnetic field (potential)	\mathbf{B}_g	—
Internal toroidal magnetic field (flux rope)	\mathbf{B}_{Ti}	—
Total poloidal magnetic field	\mathbf{B}_P	$\mathbf{B}_s + \mathbf{B}_{Pi}$
Total toroidal magnetic field	\mathbf{B}_T	$\mathbf{B}_g + \mathbf{B}_{Ti}$
Toroidal current density	\mathbf{J}_T	$\nabla \times \mathbf{B}_{Pi}/\mu_0$
Poloidal current density	\mathbf{J}_P	$\nabla \times \mathbf{B}_{Ti}/\mu_0$
Hoop force density (upward)	f_h	$\hat{\mathbf{e}}_z \cdot (\mathbf{J}_T \times \mathbf{B}_{Pi})$
Strapping force density (downward)	f_s	$\hat{\mathbf{e}}_z \cdot (\mathbf{J}_T \times \mathbf{B}_s)$
Tension force density (downward)	f_t	$\hat{\mathbf{e}}_z \cdot (\mathbf{J}_P \times \mathbf{B}_T)$

This decomposition is chosen so that the quantities can be grouped into those related to the poloidal magnetic field (\mathbf{B}_s , \mathbf{B}_{Pi} , \mathbf{J}_T , f_h and f_s) and those related to the toroidal magnetic field (\mathbf{B}_g , \mathbf{B}_{Ti} , \mathbf{J}_P and f_t). The force densities, f_i are integrated to force per unit length, F_i before being compared (see Methods). Note that for simplicity, scalar representations of the vector components of \mathbf{B} and \mathbf{J} are used in the main text (for example, $B_T \equiv \hat{\mathbf{e}}_T \cdot \mathbf{B}_T$).

Quantum superposition at the half-metre scale

T. Kovachy¹, P. Asenbaum¹, C. Overstreet¹, C. A. Donnelly¹, S. M. Dickerson¹, A. Sugarbaker¹, J. M. Hogan¹ & M. A. Kasevich¹

The quantum superposition principle allows massive particles to be delocalized over distant positions. Though quantum mechanics has proved adept at describing the microscopic world, quantum superposition runs counter to intuitive conceptions of reality and locality when extended to the macroscopic scale¹, as exemplified by the thought experiment of Schrödinger's cat². Matter-wave interferometers³, which split and recombine wave packets in order to observe interference, provide a way to probe the superposition principle on macroscopic scales⁴ and explore the transition to classical physics⁵. In such experiments, large wave-packet separation is impeded by the need for long interaction times and large momentum beam splitters, which cause susceptibility to dephasing and decoherence¹. Here we use light-pulse atom interferometry^{6,7} to realize quantum interference with wave packets separated by up to 54 centimetres on a timescale of 1 second. These results push quantum superposition into a new macroscopic regime, demonstrating that quantum superposition remains possible at the distances and timescales of everyday life. The sub-nanokelvin temperatures of the atoms and a compensation of transverse optical forces enable a large separation while maintaining an interference contrast of 28 per cent. In addition to testing the superposition principle in a new regime, large quantum superposition states are vital to exploring gravity with atom interferometers in greater detail. We anticipate that these states could be used to increase sensitivity in tests of the equivalence principle^{8–12}, measure the gravitational Aharonov–Bohm effect¹³, and eventually detect gravitational waves¹⁴ and phase shifts associated with general relativity¹².

Progress in the ability to manipulate quantum systems has enabled experimental tests of the foundations of quantum mechanics. These include studies of entanglement¹⁵, tests of local realism with Bell experiments^{16,17}, and exploration of wave–particle duality in delayed choice experiments with photons¹⁸ and atoms¹⁹. The quantum superposition principle is a central axiom of quantum mechanics, and efforts to test its universal validity have attracted much interest¹. A breakdown of quantum superposition at large scales could arise from fundamental modifications to quantum dynamics^{4,5}, interaction with a field of cosmological origin⁵, or quantum gravitational effects^{1,5}. Currently, the best bounds on such decoherence mechanisms at large length scales come from matter-wave interference experiments^{1,4}. No violations of the quantum superposition principle have yet been detected. To bound or discover such violations at macroscopic scales requires a well-controlled system that limits dephasing and decoherence from conventional and technical sources.

Atom interferometry offers a way to create and characterize atomic superpositions. The field of atom interferometry has developed as a long series of experiments originating from Bordé's realization of the importance of recoil effects in precision Ramsey laser spectroscopy^{6,20}, which led to the Bordé–Ramsey technique^{6,20}. Other important developments include the demonstration of atom interferometers using mechanical gratings²¹ and two-photon transitions⁷.

To create large atomic quantum superpositions, a significant challenge is to combine large momentum transfer (LMT) atomic beam splitters^{22,23} with long-time (>2 s) atom interferometry^{24,25}.

Interferometers with LMT beam splitters are susceptible to dephasing from laser intensity inhomogeneity and wavefront perturbations across the atom cloud. These dephasing mechanisms are coupled to the transverse expansion of the atom cloud and are therefore exacerbated by long interferometer durations.

We achieve long free-fall times by launching a Bose–Einstein condensed cloud of $\sim 10^5$ ultracold ^{87}Rb atoms into a 10 m atomic fountain using a chirped optical lattice²⁴. After the lattice launch, we use a sequence of optical pulses to apply a beam splitter that places each atom into a superposition of two wave packets with different momenta, corresponding to the two arms of a Mach–Zehnder interferometer⁷. We then allow the two wave packets to spatially separate vertically during a drift time $T = 1.04$ s. Subsequently, we redirect the two wave packets back towards each other with additional optical pulses (the mirror sequence), and cause them to interfere using a final beam splitter when they once again spatially overlap after another drift interval of $T = 1.04$ s. Finally, we image the two interferometer output ports using a CCD camera (see Fig. 1).

The maximum spatial separation reached in the interferometer is $\Delta z = n(\hbar k/m)T$, where k is the laser wave number, n is the number of photon recoils ($\hbar k$) transferred by the beam splitter, and m is the atomic mass ($\hbar k/m$ is the velocity associated with a single photon momentum recoil). Our LMT beam splitters transfer up to $90\hbar k$, yielding superpositions with much larger spatial separation than is possible with conventional $2\hbar k$ atom optics (54 cm for $90\hbar k$, as shown in Fig. 2). We realize the beam splitters with sequential $2\hbar k$ Bragg transitions²³ (see Methods). The laser beams that drive the Bragg transitions are sent into the atomic fountain from the top and retroreflected by a mirror at the bottom.

To quantify the coherence of the macroscopic superposition states, we measure the contrast of the interferometer. To determine the contrast, we record the amount of variation in the normalized population in one of the output ports as it varies between constructive and destructive interference. The normalized population in output port i is $P_i \equiv N_i/(N_1 + N_2)$, where N_i is the measured atom number in output port i . Owing to interference between the two arms of the interferometer, the population oscillates between the two output ports²⁴. Examples of fluorescence images showing this population modulation are given in Fig. 3.

Owing to the large enclosed space-time area ΔzT , the interferometer is highly sensitive to acceleration. Specifically, the sensitivity of the interferometer phase ϕ to an acceleration a can be expressed as⁷ $\Delta\phi = ma\Delta zT/\hbar$. This leads to an acceleration response for our interferometer of 2×10^8 rad per g for $2\hbar k$ beam splitters and 8×10^9 rad per g for $90\hbar k$ beam splitters (g is the acceleration due to gravity). Consequently, the interferometer phase fluctuates by much more than 2π from shot to shot due to vibration of the retroreflection mirror, causing the output ports to vary randomly between constructive and destructive interference. Therefore, we see significant contrast, but the large acceleration sensitivity prevents the observation of a stable fringe as the phase is scanned. Since the contrast quantifies the coherence of the macroscopic superposition states, the contrast is the relevant metric for this work (as in photon recoil measurements with contrast

¹Department of Physics, Stanford University, Stanford, California 94305, USA.

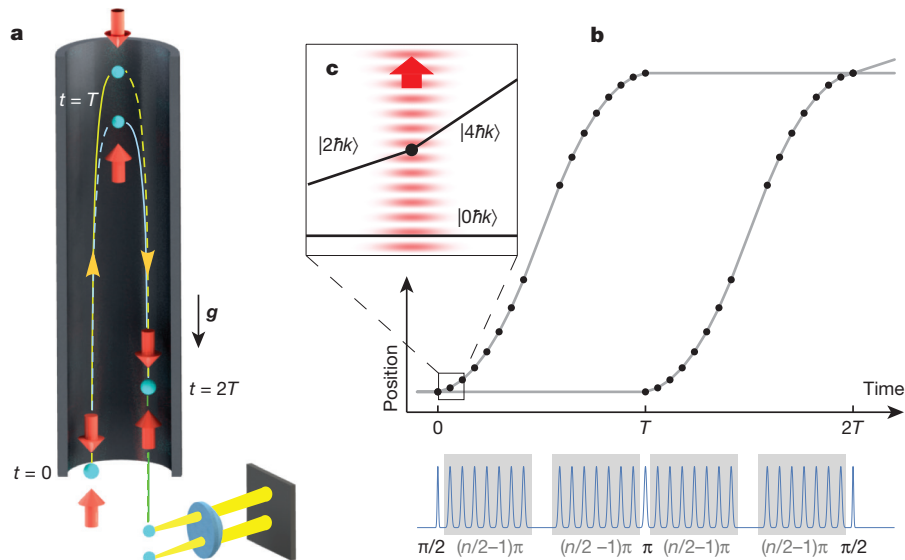


Figure 1 | Fountain interferometer. **a**, After evaporative cooling and a magnetic lensing sequence (see Methods), the ultra-cold atom cloud is launched vertically from below the cylindrical magnetic shield using an optical lattice. At $t = 0$, the first beam splitter sequence splits the cloud into a superposition of momentum states separated by $n\hbar k$. At $t = T$, the wave packet is fully separated, and a mirror sequence reverses the momentum states of the two halves of the cloud. At $t = 2T$, the clouds spatially overlap, and a final beam splitter sequence is applied. After a short drift time, the output ports spatially separate by 6 mm owing to their differing momenta, and the two complementary ports are imaged. This diagram is not to scale, and the upward- and downward-going clouds are shown horizontally displaced for clarity. The red, cylindrical arrows illustrate the counter-propagating laser beams that drive the Bragg transitions. The blue spheres

represent the atomic wave packets. The solid and dashed lines show the trajectories of the atomic wave packets (solid lines correspond to $n\hbar k$ greater momentum in the upward direction than the dashed lines), and the yellow arrowheads indicate the direction of motion. **b**, Pulse sequence of a $16\hbar k$ interferometer, see Methods for details. The main plot depicts the spacetime trajectories of the wave packets, and the pulse train underneath shows the temporal profile of the laser pulse sequences. **c**, A moving standing wave (red wave, direction of motion indicated by red arrow) induces a Bragg transition of one specific velocity class and changes its momentum by $2\hbar k$, for example, from $2\hbar k$ to $4\hbar k$. The black lines show a zoomed-in view of the spacetime trajectories, labelled by momentum. The black dot indicates the point at which the transition from momentum $2\hbar k$ to $4\hbar k$ occurs.

interferometry²⁶). In many future experiments to explore gravitational physics, differential measurement schemes²⁷ (for example, gravity gradiometry) will be used to exploit the increased sensitivity offered by large superposition states while cancelling the vibration-induced phase noise as a common mode^{12–14}. In the work presented here, common-mode cancellation of the vibration-induced phase noise between different parts of the atom cloud allows us to observe contrast and additionally to see spatial interference fringes across the atom cloud (see below).

To further demonstrate interference, we measure the contrast envelope, that is, the variation of P_1 as a function of a timing delay δT before the final beam-recombining pulse sequence. At suitably large delays, contrast is suppressed, thus allowing characterization of technical noise sources which might be conflated with contrast at shorter delays. The timing asymmetry leads to a phase shift $nk v_z \delta T$ that depends on the vertical velocity v_z (refs 24, 25). Integrating over the vertical velocity distribution of the atom cloud after the interferometer (r.m.s. width Δv_z), the contrast is expected to decay with δT as the envelope

function²⁸ $\Gamma(\delta T) \equiv \exp[-n^2 k^2 \Delta v_z^2 \delta T^2 / 2] = \exp[-\delta T^2 / 2\delta T_c^2]$, where the coherence time is given by $\delta T_c \equiv 1/(nk \Delta v_z)$. Figure 4a displays the contrast envelopes and comparison to theory for $30\hbar k$, $60\hbar k$, and $90\hbar k$ beam splitters. We plot $\sigma(P_1)$, the standard deviation of the set of observed P_1 values after a sequence of 20 shots at the specified δT , as δT is varied (see also Extended Data Fig. 2). Note that $2\sqrt{2}\sigma(P_1)$ is approximately equal to the contrast²². The data closely match the expected decay dependence $\Gamma(\delta T)$ for the known values of n , k and Δv_z .

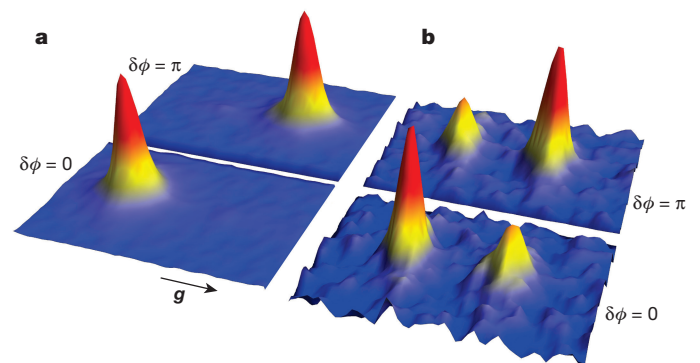


Figure 3 | Fluorescence images of output ports. The two atom clouds resulting from the final beam splitter constitute the output ports of the interferometer. A single fluorescence image allows us to extract the atom number in each port. **a**, The $2\hbar k$ interferometer shows high contrast with nearly full population oscillation between the upper port (front image) and the lower port (back image). **b**, For the $90\hbar k$ interferometer, the population oscillates by more than 40%. Owing to spontaneous emission and velocity selectivity, the detected atom number is more than ten times smaller than for $2\hbar k$. All displayed images are normalized to have the same peak height and are labelled with $\delta\phi$ corresponding to the interferometer phase modulo 2π . Each image is 13.8×9.7 mm, and the data are smoothed with a Gaussian filter with radius 0.5 mm.

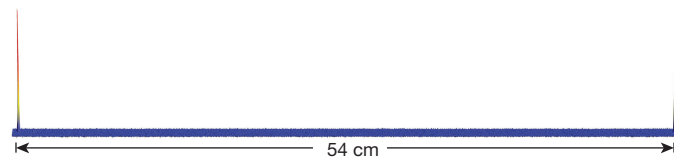


Figure 2 | Wave packets separated by 54 cm. We adjust the launch height of the millimetre-sized atom cloud so that it passes the detector when the wave packets (corresponding to the two peaks in the image) are maximally separated. In order to visualize the full extent of the wave function, we take 36 snapshots of different slices of the distribution. The images are taken at slightly different times between the atom launch and the fluorescence imaging and are stitched together according to the velocity of the atoms. The vertical height in the plot corresponds to atom density (red indicates higher density).

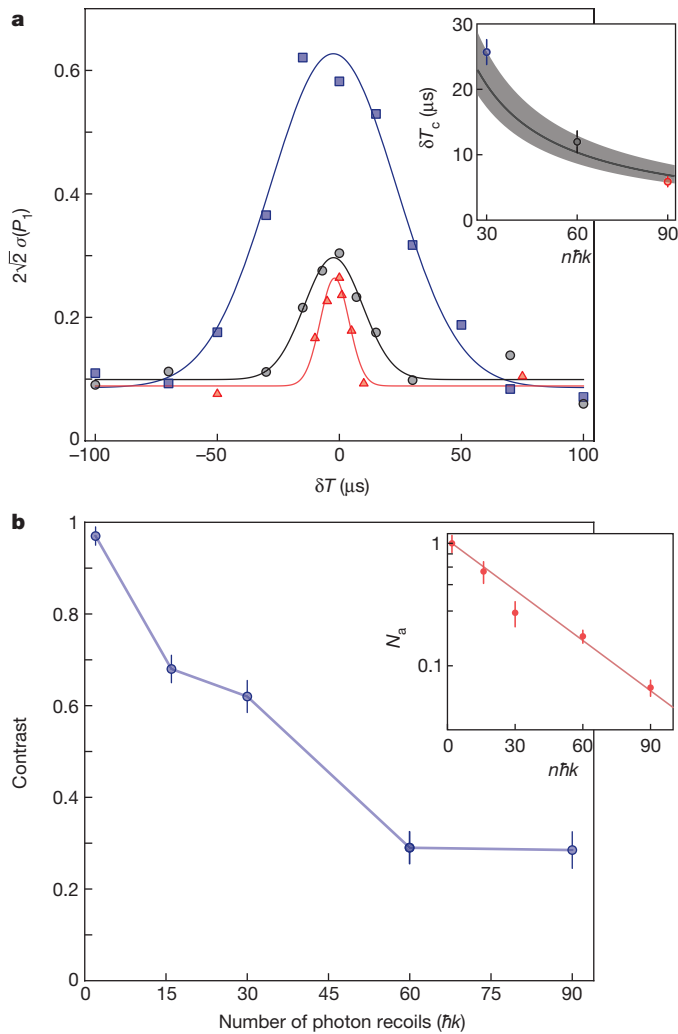


Figure 4 | Contrast metrics. **a**, The contrast envelopes establish the interference effect. We plot $2\sqrt{2}\sigma(P_1)$ versus the timing delay δT , where $\sigma(P_1)$ is the standard deviation of the set of observed P_1 values after a sequence of 20 shots at the specified δT (P_1 is the normalized population in output port 1). The data points corresponding to the blue squares, black circles and red triangles are for $30\hbar k$, $60\hbar k$ and $90\hbar k$. The solid curves show the theory $A + B\exp(\delta T - \delta T_0)$, with coherence time δT_c , offset A , centre δT_0 , and amplitude B as fitting parameters. Examples of the traces that lead to the points in the contrast envelopes are shown in Extended Data Fig. 2. Inset, comparison of fitted coherence times (points, 1 s.d. error bars from fit uncertainty) to theory (grey curve). The grey, shaded region indicates 1 s.d. theoretical uncertainty arising from uncertainty in the measured velocity spread Δv_z . **b**, Trends in maximum observed contrast (blue data points, main panel) and normalized atom number N_a in the output ports (red data points, inset) with $n\hbar k$. The data points are for $n = 2, 16, 30, 60$ and 90 . The atom number is normalized to the average number of atoms after a $2\hbar k$ interferometer. The thin, red curve in the inset shows the predicted atom number based on the measured spontaneous emission loss rate and π -pulse velocity selectivity. Error bars, 1 s.d. uncertainties computed with the analysis discussed in Methods.

Given that the atom cloud has a known time $t_e = 2.6$ s to expand, the vertical size of the interferometer output ports provides us with an independent measurement of $\Delta v_z = 0.20 \pm 0.04$ mm s⁻¹. The measured coherence times, as determined by fits of the contrast envelope widths, show quantitative agreement with their theoretically predicted values (see Fig. 4a).

Figure 4b shows the interference contrast for various values of n . To determine the contrast value for a given n , we use maximum likelihood estimation on the data corresponding to the highest point in the contrast envelope (see Methods). The model used to estimate the contrast

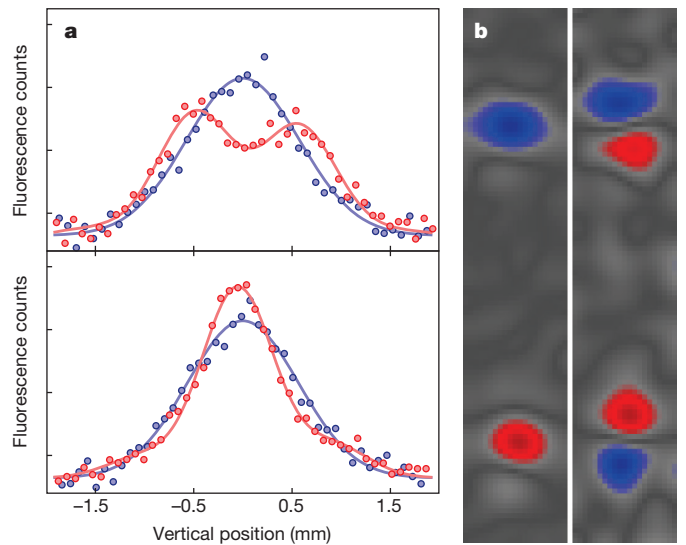


Figure 5 | Spatial interference fringes. **a**, Horizontally integrated fluorescence images of the two $30\hbar k$ output ports (upper and lower panel) for a single run with $\delta T = -50$ μ s (red). The images are fitted to a sinusoidally modulated Gaussian profile. For comparison, the output ports for $\delta T = 100$ μ s have a Gaussian profile without interference fringes (blue). y axis in arbitrary units. **b**, Cosine (left panel) and sine (right panel) principal components of a set of $30\hbar k$ interferometer runs with $\delta T = -50$ μ s, which show the effects of a vertical phase gradient across the cloud. All observed fringes are linear combinations of these basis images. Red and blue regions are anti-correlated.

corrects for the technical noise measured away from the contrast peak (that is, at large δT). Also, Fig. 4b inset shows the exponential scaling of atom loss with n . Atom loss derives from two factors: spontaneous emission decay with $1/e$ point $n = 75 \pm 10$, and residual velocity selection of the π -pulses.

A complementary demonstration of interference is the observation of spatial interference fringes across the atom cloud for small timing delays δT (refs 25,29). The predicted fringe wavelength is $\lambda_z = 2\pi t_e / (nk|\delta T - \delta T_0|)$, where t_e is the cloud expansion time and δT_0 accounts for velocity-dependent phase shifts from force gradients²⁹ (see Methods). Figure 5a shows an unsmoothed example of the directly observed fringe from a single shot. The 1σ uncertainty in the phase extracted from fitting the fringe is 0.1 rad, which is near the atom shot noise limit for the observed contrast. For $\delta T = -50$ μ s the fitted wavelength $\lambda_z = 1.5 \pm 0.1$ mm (1σ error from fit uncertainty) agrees with the theoretical value of $\lambda_z = 1.4$ mm (taking $\delta T_0 = 0$). Assuming a spherical Earth's gravity gradient would shift the prediction to $\lambda_z = 1.5$ mm. This is equivalent to $\delta T_0 = -3.5$ μ s, which is likely to be the reason why δT_0 is slightly negative for the contrast envelopes in Fig. 4a. While the overall position of the spatial fringes varies from shot to shot, the fringes on the two ports always have complementary phases, as expected. Using principal component analysis on a set of 20 images, we extract the two orthogonal modes describing the spatial fringe²⁴ (Fig. 5b).

Even for the 54 cm delocalization and a total of 180 applied optical Bragg pulses, we observe a contrast of 28%. We attribute the ability to maintain this level of contrast to two factors: the low temperature of the atoms and an absolute light shift compensation technique (see Methods). The ultra-cold cloud remains smaller than 1 mm throughout the interferometer. This reduces the contrast loss due to larger-scale inhomogeneities in laser intensity and wavefront (for example, from the 2 cm laser radial waist). The small cloud also minimizes pollution of the output ports by non-interfering atoms originating from spontaneous emission and imperfect transfer efficiency. The importance of absolute light shift compensation is demonstrated by the fact that operating without compensation almost fully eliminates the contrast for a $30\hbar k$ interferometer (see Extended Data Fig. 1). Further improvement of the contrast at large $n\hbar k$ is likely to require reduction of wavefront

perturbations, since these are intrinsically imprinted on the cloud at each pulse.

We probe the quantum superposition principle in an unprecedented regime. Extended Data Table 1 compares the wave-packet separation, interferometer duration, and mass of our superposition states to those of other matter-wave interferometers, showing that we occupy a new region of large wave-packet separation and long time. As a result, we set new bounds on macroscopic extensions of quantum mechanics (see Extended Data Fig. 3 and Methods) that introduce a decoherence mechanism for superpositions larger than a certain critical size (the critical size is a free parameter of the theory)⁴. For instance, as shown in Extended Data Fig. 3, our bound on the decoherence rate for critical sizes $\gtrsim 1$ m is 10^4 times stronger than those placed by other experiments. In addition, these large superposition states pave the way for a new generation of fundamental physics tests using ultra-sensitive atom interferometers^{12–14}. The wave-packet delocalization and coherence time demonstrated here already meet the requirements for certain proposed atomic gravitational wave detectors¹⁴. The demonstrated enclosed space-time area combined with optical atomic clock states could also enable the study of decoherence induced by general relativistic proper time³⁰.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 June; accepted 9 October 2015.

- Arndt, M. & Hornberger, K. Testing the limits of quantum mechanical superpositions. *Nature Phys.* **10**, 271–277 (2014).
- Schrödinger, E. Die gegenwärtige Situation in der Quantenmechanik. *Naturwissenschaften* **23**, 807–812 (1935).
- Cronin, A. D., Schmiedmayer, J. & Pritchard, D. E. Optics and interferometry with atoms and molecules. *Rev. Mod. Phys.* **81**, 1051–1129 (2009).
- Nimmrichter, S. & Hornberger, K. Macroscopicity of mechanical quantum superposition states. *Phys. Rev. Lett.* **110**, 160403 (2013).
- Bassi, A., Lochan, K., Satin, S., Singh, T. & Ulbricht, H. Models of wave-function collapse, underlying theories, and experimental tests. *Rev. Mod. Phys.* **85**, 471–527 (2013).
- Bordé, C. J. Atomic interferometry with internal state labelling. *Phys. Lett. A* **140**, 10–12 (1989).
- Kasevich, M. & Chu, S. Atomic interferometry using stimulated Raman transitions. *Phys. Rev. Lett.* **67**, 181–184 (1991).
- Bonnin, A., Zahzam, N., Bidet, Y. & Bresson, A. Simultaneous dual-species matter-wave accelerometer. *Phys. Rev. A* **88**, 043615 (2013).
- Schlippert, D. *et al.* Quantum test of the universality of free fall. *Phys. Rev. Lett.* **112**, 203002 (2014).
- Kuhn, C. C. N. *et al.* A Bose-condensed, simultaneous dual-species Mach-Zehnder atom interferometer. *New J. Phys.* **16**, 073035 (2014).
- Geiger, R. *et al.* Detecting inertial effects with airborne matter-wave interferometry. *Nature Commun.* **2**, 474 (2011).
- Dimopoulos, S., Graham, P., Hogan, J. & Kasevich, M. General relativistic effects in atom interferometry. *Phys. Rev. D* **78**, 042003 (2008).
- Hohensee, M. A., Estey, B., Hamilton, P., Zeilinger, A. & Müller, H. Force-free gravitational redshift: proposed gravitational Aharonov-Bohm experiment. *Phys. Rev. Lett.* **108**, 230404 (2012).
- Dimopoulos, S., Graham, P., Hogan, J., Kasevich, M. & Rajendran, S. Atomic gravitational wave interferometric sensor. *Phys. Rev. D* **78**, 122002 (2008).
- Julsgaard, B., Kozhekin, A. & Polzik, E. S. Experimental long-lived entanglement of two macroscopic objects. *Nature* **413**, 400–403 (2001).
- Aspect, A., Dalibard, J. & Roger, G. Experimental test of Bell's inequalities using time-varying analyzers. *Phys. Rev. Lett.* **49**, 1804–1807 (1982).
- Giustina, M. *et al.* Bell violation using entangled photons without the fair-sampling assumption. *Nature* **497**, 227–230 (2013).
- Jacques, V. *et al.* Experimental realization of Wheeler's delayed-choice gedanken experiment. *Science* **315**, 966–968 (2007).
- Manning, A. G., Khakimov, R. I., Dall, R. G. & Truscott, A. G. Wheeler's delayed-choice gedanken experiment with a single atom. *Nature Phys.* **11**, 539–542 (2015).
- Bordé, C. J. *et al.* Observation of optical Ramsey fringes in the 10 μ m spectral region using a supersonic beam of SF₆. *J. Phys. (Paris)* **42**(C8), 15–19 (1981).
- Keith, D. W., Ekstrom, C. R., Turchette, Q. A. & Pritchard, D. E. An interferometer for atoms. *Phys. Rev. Lett.* **66**, 2693–2696 (1991).
- Müller, H., Chiow, S.-W., Long, Q., Herrmann, S. & Chu, S. Atom interferometry with up to 24-photon-momentum-transfer beam splitters. *Phys. Rev. Lett.* **100**, 180405 (2008).
- Chiow, S.-W., Kovachy, T., Chien, H.-C. & Kasevich, M. A. $102\hbar k$ large area atom interferometers. *Phys. Rev. Lett.* **107**, 130403 (2011).
- Dickerson, S. M., Hogan, J. M., Sugarbaker, A., Johnson, D. M. S. & Kasevich, M. A. Multiaxis inertial sensing with long-time point source atom interferometry. *Phys. Rev. Lett.* **111**, 083001 (2013).
- Müntinga, H. *et al.* Interferometry with Bose-Einstein condensates in microgravity. *Phys. Rev. Lett.* **110**, 093602 (2013).
- Gupta, S., Dieckmann, K., Hadzibabic, Z. & Pritchard, D. E. Contrast interferometry using Bose-Einstein condensates to measure \hbar/m and α . *Phys. Rev. Lett.* **89**, 140401 (2002).
- Foster, G. T., Fixler, J. B., McGuirk, J. M. & Kasevich, M. A. Method of phase extraction between coupled atom interferometers using ellipse-specific fitting. *Opt. Lett.* **27**, 951–953 (2002).
- Parazzoli, L. P., Hankin, A. M. & Biedermann, G. W. Observation of free-space single-atom matter wave interference. *Phys. Rev. Lett.* **109**, 230401 (2012).
- Sugarbaker, A., Dickerson, S. M., Hogan, J. M., Johnson, D. M. S. & Kasevich, M. A. Enhanced atom interferometer readout through the application of phase shear. *Phys. Rev. Lett.* **111**, 113002 (2013).
- Zych, M., Costa, F., Pikovski, I. & Brukner, C. Quantum interferometric visibility as a witness of general relativistic proper time. *Nature Commun.* **2**, 505 (2011).

Acknowledgements We thank S.-w. Chiow for discussions and contributions to the apparatus. T.K. acknowledges support from the Fannie and John Hertz Foundation, T.K. and C.A.D. from the NSF GRFP, and T.K. and C.O. from the Stanford Graduate Fellowship. This work was supported in part by NASA GSFC grant no. NNX11AM31A.

Author Contributions T.K., P.A., C.O., C.A.D., J.M.H. and M.A.K. carried out the experiment, analysed the data and prepared the manuscript. S.M.D. and A.S. contributed significantly to the early stages of the experiment.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.A.K. (kasevich@stanford.edu).

METHODS

Atom source. A 2D magneto optical trap (MOT) loads a 3D MOT in the centre of our 10 m vacuum tube for 4 s. We evaporate the ^{87}Rb atoms in a time-orbiting potential (TOP) trap for 14 s and apply a magnetic lensing sequence to further reduce their kinetic energy³¹. The ultra-cold atoms are then launched upwards into the interferometer region with a chirped optical lattice. Overall, we have a cycle time of roughly 22 s.

Atom optics. For the initial beam splitter, a $\pi/2$ -pulse splits the interferometer arms in momentum space by $2\hbar k$, followed by a sequence of $(n/2) - 1$ π -pulses that selectively accelerate one of the arms to increase the momentum splitting to $n\hbar k$. The mirror sequence consists of $n - 1$ sequential π -pulses that interchange the momenta of the two interferometer arms^{23,32}, and the final beam splitter sequence once again contains $(n/2) - 1$ π -pulses applied to one arm followed by a $\pi/2$ -pulse.

Bragg transitions couple different momentum states of the hyperfine level $F = 2$, $m_F = 0$. In contrast to Raman transitions^{7,33}, a Bragg scheme does not suffer from light-shift-induced variations of the hyperfine splitting between $F = 1$ and $F = 2$. The optical pulses that drive the Bragg transitions have Gaussian temporal profiles with full-width at half-maximum (FWHM) $60\ \mu\text{s}$ for π -pulses and $30\ \mu\text{s}$ for $\pi/2$ -pulses. Before the first beam splitter, the vertical velocity width is filtered by a $300\ \mu\text{s}$ π -pulse that transfers only a narrow velocity slice. The two atom optics laser beams each contain 3 W of power 30 GHz detuned from the excited state resonance and are generated by frequency doubling the outputs of 1,560 nm fibre amplifiers in nonlinear crystals³⁴. These beams are combined on a polarizing beam splitter and enter the atomic fountain from the top. They have a radial waist of 2 cm and are retroreflected by a mirror at the bottom of the fountain. The mirror's angle is adjusted between pulse sequences by a piezo-actuated tip-tilt stage to compensate for Coriolis forces from Earth's rotation²⁴. Given that the laser intensity is limited by the large beam waist, sequential $2\hbar k$ Bragg transitions offer lower spontaneous emission losses than higher order Bragg transitions³⁵.

Absolute light shift compensation. We implement a technique to compensate optical dipole forces on the atoms from imperfections in the laser beam profile. Dipole forces arise from gradients in the laser intensity, since the energy of an atomic state is shifted by an amount proportional to the local laser intensity (light shift)³. These forces can distort the cloud and cause large differential phase shifts across the cloud. The differential phase shifts occur because the laser intensity profile varies with vertical position and is therefore not fully common to the two interferometer arms. To perform this compensation, we adjust the laser spectrum so that the absolute light shift from the blue-detuned spectral content, including the frequency components that drive the Bragg transitions, is cancelled by the absolute light shift from the red-detuned spectral content.

We achieve a light-shift-compensating spectrum by phase modulating each of the two atom optics lasers at 30 GHz, with the carrier 3.4 GHz blue-detuned from resonance and nearly fully suppressed. The two atom optics lasers are offset by an AOM shift of 160 MHz so that only one pair of sidebands drives Bragg transitions. The phase modulation occurs on the 1,560 nm light seeding the fibre amplifiers. To tune the asymmetry between the red and blue sidebands, we adjust the temperature of the frequency doubling crystals. We measure the optical spectrum with a scanning Fabry–Perot cavity.

Contrast metrics data analysis. Following similar analysis from previous work¹¹, we model P_1 as a random variable. Our model for the probability density function (PDF) of P_1 includes additive Gaussian noise¹¹. P_1 is related to the phase Φ and contrast c of the interferometer by:

$$P_1(\Phi, X; c, w) = \frac{1}{2} + \frac{c}{2} \cos \Phi + X(w) \quad (1)$$

We assume that the interferometer phase is uniformly distributed, so the PDF of Φ is given by $f_\Phi(\phi) = \frac{1}{\pi}$ where $\phi \in [0, \pi]$, and that the amplitude noise X is normally distributed with standard deviation w . We also assume that Φ and X are independent, so the PDF of P_1 in the presence of noise X is equal to the convolution of the PDF of P_1 in the absence of noise ($w \rightarrow 0$) with the PDF of X .

Since the contrast approaches zero for large δT , all remaining fluctuations in P_1 at large δT are due to amplitude noise. Therefore, we estimate w by computing the standard deviation of data taken at large values of δT . To estimate c , we use maximum likelihood estimation³⁶ on the data set corresponding to the highest point in each contrast envelope, taking w to be a fixed parameter. The resulting contrast estimates are plotted in Fig. 4b. To calculate the uncertainty in the contrast estimates, we use the observed Fisher information for each data set³⁶. We also propagate the uncertainty in the measured value of w . We discuss this contrast estimation procedure in greater detail below.

Error bars for the atom number in Fig. 4b are computed from statistical standard deviation. The curve showing the predicted atom number in Fig. 4b accounts for atom loss due to spontaneous emission and imperfect π -pulse transfer efficiency. We measure the spontaneous emission loss rate by illuminating the launched cloud with a detuned interferometer pulse sequence. Specifically, all pulses are detuned from their respective two-photon resonances so that there is no transfer. Therefore, the ratio of the number of atoms remaining after such a pulse sequence to the number of atoms remaining after a launch with no pulses allows us to determine the fraction of the atoms lost due to spontaneous emission. To measure the π -pulse transfer efficiency, we apply a $\pi/2$ -pulse followed by 44 π -pulses and compare the number of atoms in the transferred peak ($90\hbar k$ total momentum kick) to the number of atoms in the peak that is left untransferred by the $\pi/2$ -pulse. Spontaneous emission loss is the same for both peaks and therefore does not confound the measurement. We note that the two peaks have the same height, while the transferred peak has a narrower vertical width (for example, see Fig. 2). This indicates that the imperfect transfer efficiency arises from π -pulse velocity selectivity.

Spatial interference fringes. Owing to the long expansion time t_e , the launched atom cloud is effectively a point source, meaning that by the time of detection the vertical velocity distribution has been mapped onto the vertical position z through the relation $z \approx v_z t_e$ (v_z is the vertical velocity). The velocity dependent phase shift $\hbar k v_z \delta T$ then leads to a position dependent phase shift²⁹ with corresponding wavelength $\lambda_z = 2\pi t_e / (\hbar k |\delta T - \delta T_0|)$. Here δT_0 accounts for any velocity-dependent phase shifts from force gradients²⁹. To observe the fringes, we reduce the fluorescence imaging time to 2.5 ms (see Fig. 5). We choose $\delta T = -50\ \mu\text{s}$ so that a full wavelength is visible on the atom cloud. For $\delta T = 100\ \mu\text{s}$ the smaller fringe period is completely blurred out by imaging heating of the atom cloud. The direct spatial interference contrast for $\delta T = -50\ \mu\text{s}$ is lower than the contrast with $\delta T = 0$ reported in Fig. 4b due to this blurring.

We use principal component analysis (PCA) to extract spatial fringes from a set of 20 interferometer runs. In addition to the fringe pattern, PCA is sensitive to shot-to-shot variation of the centre of mass position of the cloud. To minimize crosstalk between these effects, we correct for the vertical and horizontal motion before performing PCA. We find the position of the cloud centre of mass for each shot using Gaussian fits and then shift each image appropriately to remove the motion. The data are also smoothed with a $400\ \mu\text{m}$ Gaussian filter before PCA. We identify the first principal component as the shape of the overall cloud envelope. Principal components two and three correspond to the cosine and sine components of the fringe pattern (Fig. 5).

Testing macroscopic extensions of quantum mechanics. In Extended Data Fig. 3, we show exclusion curves for the parameter space of a general class of minimal modifications to quantum mechanics⁴. The theory is characterized by two parameters: a critical length scale \hbar/σ_q beyond which quantum superpositions decay (σ_q corresponds to the magnitude of spontaneous momentum kicks introduced by the modification), and a survival time τ_e that it takes for this decay to happen for an electron superposition larger than \hbar/σ_q . Therefore, different experiments can be referenced to an electron for comparison⁴. The critical length scale and survival time are free parameters of the theory that must be determined by experiment—there is no a priori assumption as to what their values should be⁴.

To compare the bounds set by our experiment to the previous experimental status quo, we include exclusion curves for a number of other matter wave interference experiments that place bounds on this parameter space: atom interferometry with rubidium²⁵, caesium^{37–39} and sodium⁴⁰; neutron interferometry⁴¹; and interferometry with large molecules^{42,43}. Molecular interferometry provides its strongest bounds on modifications to quantum mechanics of this form for submicrometre critical length scales, whereas the bounds from atom interferometry dominate at larger critical length scales due to the large wave packet separation.

We note that there are experiments demonstrating the preservation of entanglement over long distances, such as Bell experiments with photons⁴⁴ and the entanglement of many atomic spins¹⁵. While these experiments test quantum mechanics in a complementary way by generating entangled states, they do not create spatial superpositions of massive particles and thus do not bound the parameter space considered here.

Interferometer noise model and contrast estimation. Following ref. 11, we model the normalized population $P_1 \equiv N_1/(N_1 + N_2)$ of an interferometer port as a random variable. P_1 is related to the phase Φ and contrast c of the interferometer by equation (1) above. We also assume that the interferometer noise X is normally distributed with PDF

$$f_X(x; w) = \frac{1}{w\sqrt{2\pi}} e^{-x^2/2w^2} \quad (2)$$

and that Φ and X are independent.

In the absence of noise ($w \rightarrow 0$), the PDF of P_1 is given by

$$g_{P_1}(p; c) = \frac{2}{\pi} \frac{1}{\sqrt{c^2 - (2p - 1)^2}}. \quad (3)$$

This function is supported on $\left(\frac{1}{2} - \frac{c}{2}, \frac{1}{2} + \frac{c}{2}\right)$ and has asymptotes at the boundaries. Since Φ and X are independent, the PDF of P_1 for non-zero w can be computed by convolving $g_{P_1}(p; c)$ with $f_X(x; w)$:

$$f_{P_1}(p; c, w) = \int_{1/2-c/2}^{1/2+c/2} g_{P_1}(\tau; c) f_X(p - \tau; w) d\tau \quad (4)$$

To experimentally determine w , we make the interferometer asymmetry δT large enough that $c \rightarrow 0$. In this case, P_1 is normally distributed, and the observed residual fluctuation in P_1 is used to estimate w . For the data reported in this work, we typically find $w \approx 0.03 \pm 0.005$. To estimate c , we use the maximum likelihood method³⁶ on a sequence of shots $\{p_1, \dots, p_m\}$ at fixed δT . Specifically, we compute the likelihood

$$L(c; w, \{p_1, \dots, p_m\}) = \prod_{i=1}^m f_{P_1}(p_i; c, w) \quad (5)$$

taking the data points p_i and the measured value of w to be fixed parameters. The most likely value of c given the data is found by maximizing L as a function of c , or equivalently by solving

$$\frac{\partial}{\partial c} \ln L = 0 \Rightarrow \frac{\partial}{\partial c} \sum_{i=1}^m \ln f_{P_1}(p_i; c, w) = 0. \quad (6)$$

We maximize L numerically to generate the contrast estimates plotted in Fig. 4b.

The uncertainty in these contrast estimates arises from two sources. First, the standard error $\sigma_c(c)$ of the maximum likelihood method scales as the square root of the inverse of the Fisher information in the limit of a large number of samples m . The Fisher information $F(c)$ is defined by

$$F(c) = \int \left(\frac{\partial}{\partial c} \ln f_{P_1}(p; c, w) \right)^2 f_{P_1}(p; c, w) dp \quad (7)$$

In the asymptotic limit $m \rightarrow \infty$, we have:

$$\sigma_c(c) = \frac{1}{\sqrt{m}} \frac{1}{\sqrt{F(c)}} \quad (8)$$

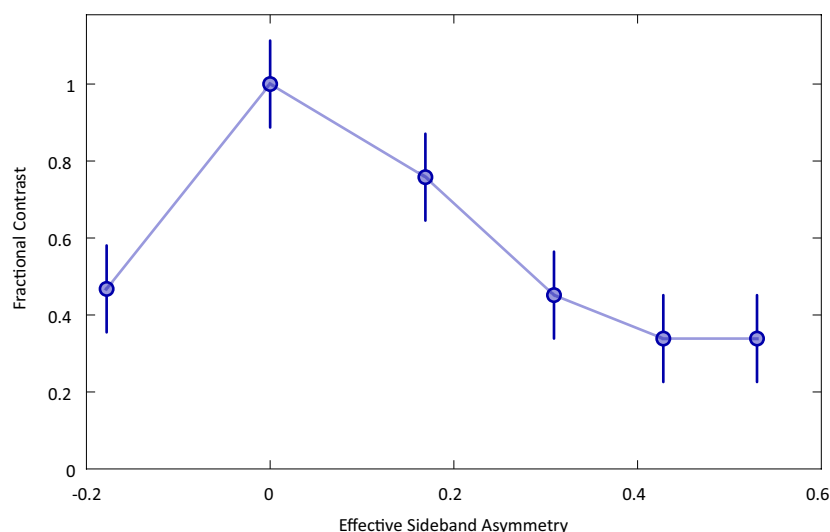
For $m \geq 20$, the error in the asymptotic approximation does not significantly contribute to the uncertainty. We verify this by computing the observed Fisher information F_o for each data set, where:

$$F_o(c; w, \{p_1, \dots, p_m\}) = -\frac{1}{m} \frac{\partial^2}{\partial c^2} \sum_{i=1}^m \ln f_{P_1}(p_i; c, w) \quad (9)$$

Second, statistical uncertainty in the measurement of w propagates into uncertainty in the estimate of c . Both of these sources of uncertainty are reflected in the error bars shown in Fig. 4b.

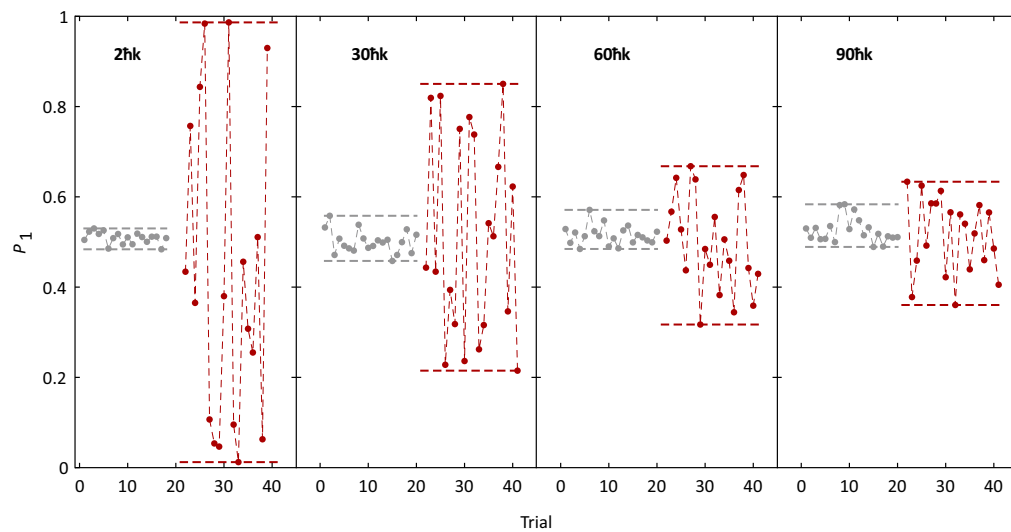
Sample size. No statistical methods were used to predetermine sample size.

31. Kovachy, T. *et al.* Matter wave lensing to picokelvin temperatures. *Phys. Rev. Lett.* **114**, 143004 (2015).
32. McGuirk, J. M., Snadden, M. J. & Kasevich, M. A. Large area light-pulse atom interferometry. *Phys. Rev. Lett.* **85**, 4498–4501 (2000).
33. Rosi, G., Sorrentino, F., Cacciapuoti, L., Prevedelli, M. & Tino, G. M. Precision measurement of the Newtonian gravitational constant using cold atoms. *Nature* **510**, 518–521 (2014).
34. Chiow, S.-W., Kovachy, T., Hogan, J. M. & Kasevich, M. A. Generation of 43 W of quasi-continuous 780 nm laser light via high-efficiency, single-pass frequency doubling in periodically poled lithium niobate crystals. *Opt. Lett.* **37**, 3861–3863 (2012).
35. Szigeti, S. S., Debs, J. E., Hope, J. J., Robins, N. P. & Close, J. D. Why momentum width matters for atom interferometry with Bragg pulses. *New J. Phys.* **14**, 023009 (2012).
36. Greene, W. H. *Econometric Analysis* (Pearson, 2012).
37. Peters, A., Chung, K. Y. & Chu, S. High-precision gravity measurements using atom interferometry. *Metrologia* **38**, 25–61 (2001).
38. Chung, K.-Y., Chiow, S.-w., Herrmann, S., Chu, S. & Müller, H. Atom interferometry tests of local Lorentz invariance in gravity and electrodynamics. *Phys. Rev. D* **80**, 016002 (2009).
39. Lan, S.-Y., Kuan, P.-C., Estey, B., Haslinger, P. & Müller, H. Influence of the Coriolis force in atom interferometry. *Phys. Rev. Lett.* **108**, 090402 (2012).
40. Kasevich, M. & Chu, S. Measurement of the gravitational acceleration of an atom with a light-pulse atom interferometer. *Appl. Phys. B* **54**, 321–332 (1992).
41. Zawisky, M., Baron, M., Loidl, R. & Rauch, H. Testing the world's largest monolithic perfect crystal neutron interferometer. *Nucl. Instrum. Meth. A* **481**, 406–413 (2002).
42. Brezger, B. *et al.* Matter-wave interferometer for large molecules. *Phys. Rev. Lett.* **88**, 100404 (2002).
43. Eibenberger, S., Gerlich, S., Arndt, M., Mayor, M. & Tüxen, J. Matter-wave interference of particles selected from a molecular library with masses exceeding 10,000 amu. *Phys. Chem. Chem. Phys.* **15**, 14696–14700 (2013).
44. Ma, X.-S. *et al.* Quantum teleportation over 143 kilometres using active feed-forward. *Nature* **489**, 269–273 (2012).



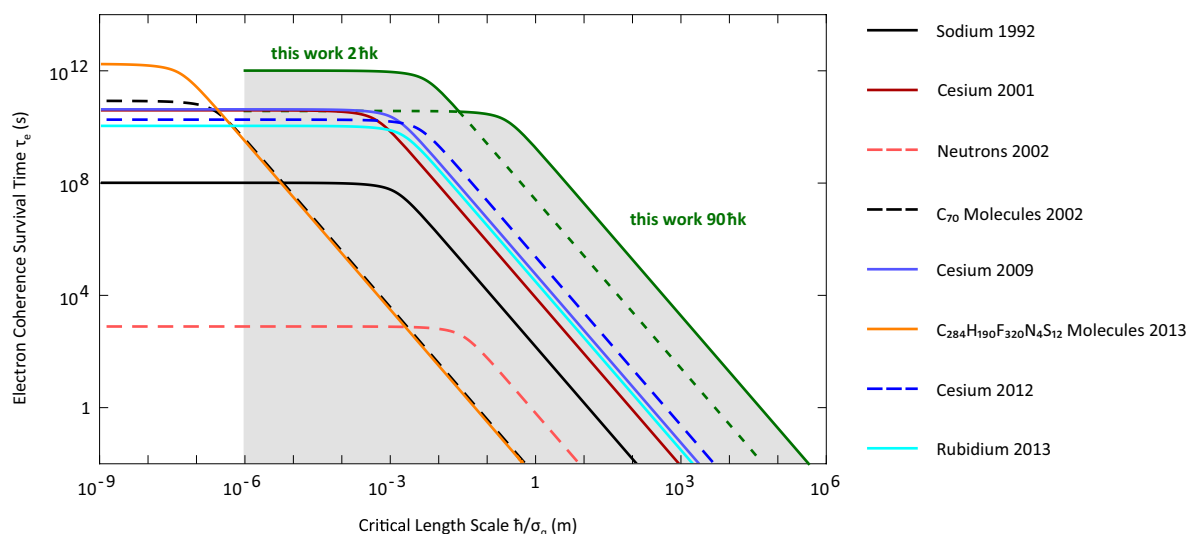
Extended Data Figure 1 | Dependence of contrast on absolute light shift compensation. For $30\hbar k$, the contrast as a fraction of its maximum value is plotted as a function of the asymmetry between the red and blue sidebands for one of the atom optics laser beams. To change the sideband asymmetry, we adjust the temperature of one of the frequency doubling crystals while keeping the sidebands of the second atom optics laser beam symmetric. Where P_{red} and P_{blue} are the respective optical powers in the red and blue sidebands, we define an asymmetry parameter $1 - (P_{\text{red}}/P_{\text{blue}})$. Since the blue sideband is used to drive the Bragg transitions, we keep P_{blue} fixed in order to maintain constant Rabi frequency. This prevents us from reaching large negative values of the asymmetry parameter, because there is only

enough total optical power available to increase P_{red} slightly without suppressing P_{blue} . In order to achieve a more negative effective value of the asymmetry parameter, we suppress the power in the carrier to half its usual amount for the one negative point in the plot. The carrier is blue detuned, so decreasing its power pulls the absolute light shift in the same direction as decreasing P_{blue} . To account for this, we plot the fractional contrast versus the effective asymmetry parameter that would yield the same light shift as the one that we implement, but at a fixed carrier power. The observed dependence of contrast on the sideband asymmetry indicates the importance of absolute light shift compensation for LMT interferometry. Error bars, 1σ .



Extended Data Figure 2 | Examples of data showing interference contrast. Plots of P_1 versus experimental trial for 2hk, 30hk, 60hk and 90hk. The red traces have small values of δT and therefore display interference contrast. As discussed in the main text, we do not observe a stable fringe because of the vibration of the retroreflection mirror. For comparison, the grey traces have large values of δT so that contrast is

eliminated, and they therefore show the amount of background amplitude noise in P_1 . Panels from left to right as follows. 2hk: red trace, $\delta T = 0 \mu\text{s}$; grey trace, $\delta T = 2 \text{ ms}$. 30hk: red trace, $\delta T = -15 \mu\text{s}$; grey trace, $\delta T = 100 \mu\text{s}$. 60hk: red trace, $\delta T = 0 \mu\text{s}$; grey trace, $\delta T = 100 \mu\text{s}$. 90hk: red trace, $\delta T = 1 \mu\text{s}$; grey trace, $\delta T = -50 \mu\text{s}$.



Extended Data Figure 3 | Bounds on macroscopic extensions of quantum mechanics. Exclusion curves for the minimal modification to quantum mechanics proposed in ref. 4. Points in this parameter space below a given curve in the plot have been ruled out by the corresponding experiment. The green curves show the bounds placed by the $2\hbar k$ and $90\hbar k$ atom interferometry results presented in this work. The grey, shaded area illustrates the region of parameter space excluded by these results. For sub-micrometre critical lengths, affected atoms would receive sufficiently large spontaneous momentum kicks to move out of the interferometer output ports. This results in atom loss and in a reduced sensitivity of the interference contrast to the decoherence rate. Therefore, we cut off

the curves arising from our interferometry data at $1\mu\text{m}$. We also show exclusion curves from a sodium interferometer from 1992⁴⁰ (solid black), a caesium interferometer from 2001³⁷ (solid red), a neutron interferometer from 2002⁴¹ (dashed red), a C_{70} molecular interferometer from 2002⁴² (dashed black), a caesium interferometer from 2009³⁸ (solid blue), a caesium interferometer from 2012³⁹ (dashed blue), a $C_{284}H_{190}F_{320}N_4S_{12}$ molecular interferometer from 2013⁴³ (solid orange), and a rubidium interferometer from 2013²⁵ (solid cyan). For all of the exclusion curves, the change in slope occurs at a critical length scale value equal to the wave packet separation.

Extended Data Table 1 | Comparison with other matter-wave interference experiments

Description	Wave packet		Mass m (amu)	Acceleration sensitivity
	separation Δz (m)	Duration T (s)		factor $m\Delta zT/\hbar$ (m/s ²) ⁻¹
This work, Rb, 90 ħk	0.54	1.04	86.9	8×10^8
Cs, 2012	9×10^{-3}	0.25	132.9	5×10^6
Cs, 2009	3×10^{-3}	0.4	132.9	3×10^6
Rb, 2013	4×10^{-3}	0.35	86.9	2×10^6
Cs, 2001	1.1×10^{-3}	0.16	132.9	4×10^5
Na, 1992	3×10^{-3}	0.05	23	5×10^4
C ₂₈₄ H ₁₉₀ F ₃₂₀ N ₄ S ₁₂ , 2013	$\sim 3 \times 10^{-7}$	1.2×10^{-3}	$\sim 10^4$	60
Neutrons, 2002	0.07	4×10^{-5}	1.01	40
C ₇₀ , 2002	$\sim 10^{-6}$	1.9×10^{-3}	840	30

We compare the wave packet separation Δz , the duration T between the beam splitter and mirror sequences, and the mass m to those of a sodium interferometer from 1992⁴⁰, a caesium interferometer from 2001³⁷, a neutron interferometer from 2002⁴¹, a C₇₀ molecular interferometer from 2002⁴², a caesium interferometer from 2009³⁸, a caesium interferometer from 2012³⁹, a C₂₈₄H₁₉₀F₃₂₀N₄S₁₂ molecular interferometer from 2013⁴³, and a rubidium interferometer from 2013²⁵. Additionally, we compare the factor $m\Delta zT/\hbar$, which is directly related to the acceleration sensitivity (see the discussion of acceleration sensitivity in the main text). The wave-packet separation in our experiment is nearly an order of magnitude larger than the next largest value (from a neutron interferometer), and the duration in our experiment is more than four orders of magnitude longer than in the neutron interferometer with a nearly hundred times larger mass.

Single-chip microprocessor that communicates directly using light

Chen Sun^{1,2*}, Mark T. Wade^{3*}, Yunsup Lee^{1*}, Jason S. Orcutt^{2†*}, Luca Alloatti², Michael S. Georgas², Andrew S. Waterman¹, Jeffrey M. Shainline^{3†}, Rimas R. Avizienis¹, Sen Lin¹, Benjamin R. Moss², Rajesh Kumar³, Fabio Pavanello³, Amir H. Atabaki², Henry M. Cook¹, Albert J. Ou¹, Jonathan C. Leu², Yu-Hsin Chen², Krste Asanović¹, Rajeev J. Ram², Miloš A. Popović³ & Vladimir M. Stojanović¹

Data transport across short electrical wires is limited by both bandwidth and power density, which creates a performance bottleneck for semiconductor microchips in modern computer systems—from mobile phones to large-scale data centres. These limitations can be overcome^{1–3} by using optical communications based on chip-scale electronic–photonic systems^{4–7} enabled by silicon-based nanophotonic devices⁸. However, combining electronics and photonics on the same chip has proved challenging, owing to microchip manufacturing conflicts between electronics and photonics. Consequently, current electronic–photonic chips^{9–11} are limited to niche manufacturing processes and include only a few optical devices alongside simple circuits. Here we report an electronic–photonic system on a single chip integrating over 70 million transistors and 850 photonic components that work together to provide logic, memory, and interconnect functions. This system is a realization of a microprocessor that uses on-chip photonic devices to directly communicate with other chips using light. To integrate electronics and photonics at the scale of a microprocessor chip, we adopt a ‘zero-change’ approach to the integration of photonics. Instead of developing a custom process to enable the fabrication of photonics¹², which would complicate or eliminate the possibility of integration with state-of-the-art transistors at large scale and at high yield, we design optical devices using a standard microelectronics foundry process that is used for modern microprocessors^{13–16}. This demonstration could represent the beginning of an era of chip-scale electronic–photonic systems with the potential to transform computing system architectures, enabling more powerful computers, from network infrastructure to data centres and supercomputers.

The electro-optic system on a chip (Fig. 1) contains a dual-core RISC-V instruction set architecture¹⁷ (ISA) microprocessor and an independent 1 MB bank of static random access memory that is used for memory. The on-chip electro-optic transmitters and receivers enable both the microprocessor and the memory to communicate directly to off-chip components using light, without the need for separate chips or components to host the optical devices. The chip was fabricated using a commercial high-performance 45-nm complementary metal–oxide semiconductor (CMOS) silicon-on-insulator (SOI) process¹⁸. No changes to the foundry process were necessary to accommodate photonics and all optical devices were designed to comply with the native process-manufacturing rules. This ‘zero-change’ integration enables high-performance transistors on the same chip as optics, reuse of all existing designs in the process, compatibility with electronics design tools, and manufacturing in an existing high-volume foundry.

The process includes a crystalline-silicon layer that is patterned to form both the body of the electronic transistors and the core

of the optical waveguides. A thin buried-oxide layer separates the crystalline-silicon layer from the silicon-handle wafer (Extended Data Fig. 1). Because the buried-oxide layer is <200 nm thick, light propagating in crystalline-silicon waveguides will evanescently leak into the silicon-handle wafer, resulting in high waveguide loss. To resolve this, we perform selective substrate removal on the chips after electrical packaging to etch away the silicon handle under regions with optical devices (Extended Data Fig. 2). We leave the silicon handle intact under the microprocessor and memory (which dissipate the most power) to allow a heat sink to be contacted, if necessary. Substrate removal has a negligible effect on the electronics¹³ and the processor is completely functional even with a fully removed substrate.

Silicon-germanium (SiGe) is present, although in low germanium mole fractions, in advanced CMOS processes to enhance hole mobility and transistor performance via compressive strain engineering of p-channel transistors¹⁸. Selecting a 1,180-nm wavelength band for the optical channel enables the use of photodetectors built using this SiGe (ref. 19). Silicon is transparent at 1,180 nm and no adverse effects are observed. At these wavelengths, the optical propagation loss in silicon-strip waveguides is 4.3 dB cm^{−1} (losses at industry-standard wavelengths of 1,300 nm and 1,550 nm are 3.7 dB cm^{−1} and 4.6 dB cm^{−1}, respectively¹³). The receiver circuit²⁰ resolves photocurrent produced by the illuminated photodetector into digital ones and zeros. The receiver sensitivity in optical modulation amplitude (OMA) is −5 dBm for a bit error ratio better than 10^{−12}.

The electro-optic transmitter consists of an electro-optic modulator and its electronic driver. The modulator is a silicon micro-ring resonator with a diameter of 10 μm, coupled to a waveguide. We dope the structure with the n-well and p-well implants used for transistors to form radially extending p–n junctions, interleaved along the azimuthal dimension^{21,22}, taking the form of a ‘spoked ring’. The ring exhibits a sharp, notched-filter optical transmission response, with a stop-band at the resonant wavelength of the ring (λ_0). Applying a negative voltage across the junctions depletes the ring of free carriers (electron and hole concentrations), while a small positive voltage refills the carriers. A change in carrier concentration influences the refractive index of the ring waveguide as a result of the carrier plasma dispersion effect²³, which, in turn, shifts λ_0 . Electro-optic modulation (on-off keying) is achieved by changing the voltage applied across the junction to move the λ_0 stop-band in and out of the laser wavelength (λ_L). The modulator has a loaded quality factor of approximately 10,000, and a voltage swing of only 1 V_{pp} (where V_{pp} is the peak-to-peak voltage) across the modulator achieves on:off ratios of 6 dB at an insertion loss of 3 dB for non-return-to-zero binary data. The low voltage, near-zero quiescent current, and low capacitance (15 fF, including wiring capacitance) result in an energy-efficient modulator driven by a standard CMOS logic

¹University of California, Berkeley, Berkeley, California 94720, USA. ²Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ³University of Colorado, Boulder, Colorado 80309, USA. [†]Present addresses: IBM T. J. Watson Research Center, Yorktown Heights, New York 10598, USA (J.S.O.); National Institute for Standards and Technology, Boulder, Colorado 80305, USA (J.M.S.).

*These authors contributed equally to this work.

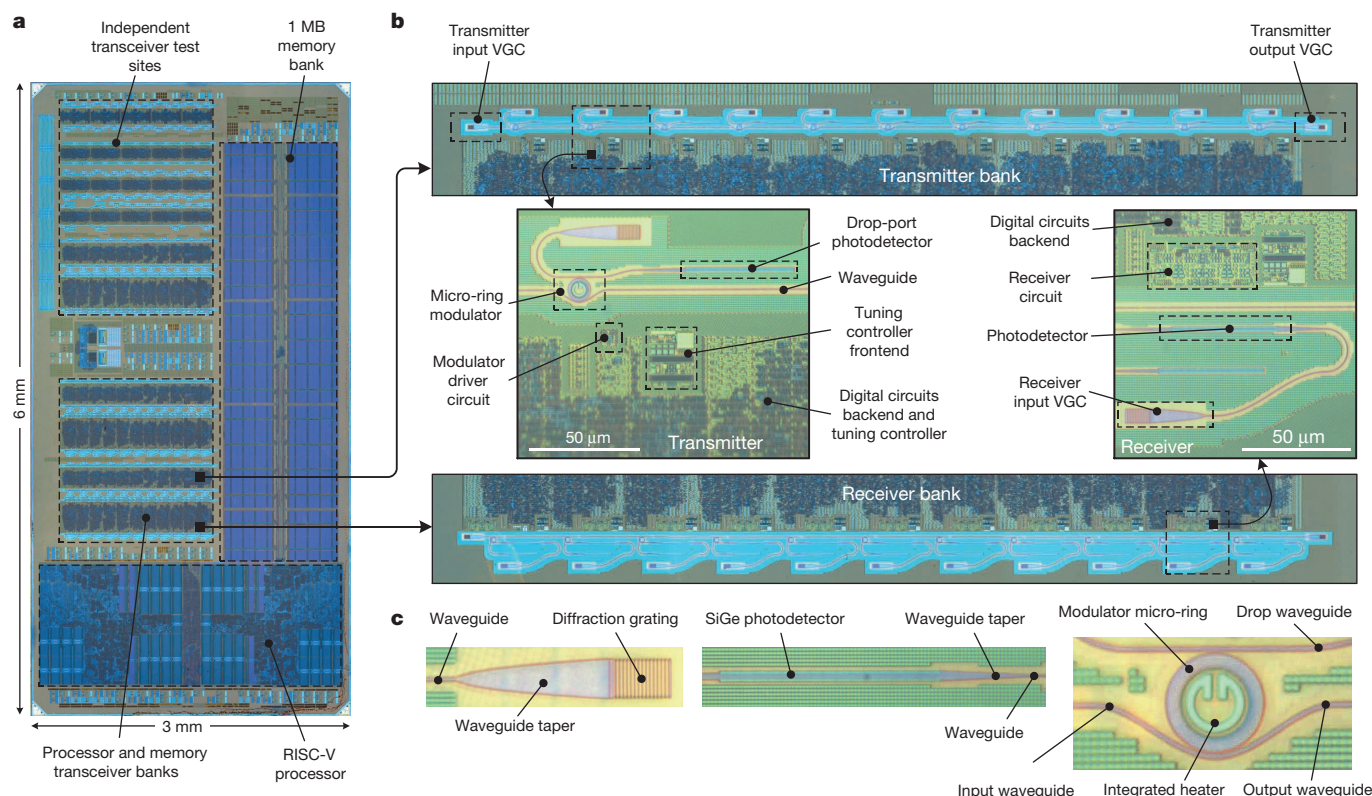


Figure 1 | The electro-optic system on a chip. **a**, Die photo of the 3 mm × 6 mm chip showing the locations and relative sizes of the processor, memory, and transceiver banks, imaged from the backside of the chip. **b**, The processor transmitter and receiver banks (the memory

transmitter and receiver banks are identical) with close-ups of individual transmitters and receivers sites. **c**, Micrographs of the grating coupler, photodetector, and resonant micro-ring modulators (left to right).

inverter at gigabit data rates using the same 1-V nominal supply that powers digital electronics.

As a resonant device, the modulator is highly sensitive to variations in the thickness of the crystalline-silicon layer within and across SOI wafers²⁴ as well as to spatially and rapidly temporally varying thermal environments created by the electrical components on the chip^{25,26}. Both effects cause λ_0 to deviate from the design value, necessitating tuning circuitry. We embedded a 400-Ω resistive microheater inside the ring to efficiently tune λ_0 and added a monitoring photodetector weakly coupled to the modulator drop port. When light resonates in the modulator ring, a small fraction of it couples to and illuminates the photodetector. This generates photocurrent proportional to the amount of resonating light, which is maximized when $\lambda_0 = \lambda_L$ (modulator is directly on resonance). Taking advantage of the densely integrated electronics, we designed a digital controller that monitors the photocurrent and controls the power to the microheater to keep λ_0 locked to λ_L under thermal variations²⁰. When λ_0 has a large offset from λ_L , such as during chip power-up, and when no photocurrent feedback is available, the controller ‘sweeps’ λ_0 by stepping the power output of the heater up or down. This sweep works to reduce the λ_0 -to- λ_L offset until sufficient photocurrent to begin the main feedback loop is obtained. The controller achieves initial lock ($\lambda_0 = \lambda_L$) within 7 ms and has a tracking time constant of 13 μs after lock-on. This system provides up to 3 nm of change in λ_0 and can compensate temperature swings of 60 K (ref. 20), aided by the superior thermal isolation afforded by selective substrate removal.

We use the direct chip-to-chip optical connectivity of the microprocessor chip to build a photonically connected main memory system for the microprocessor (Fig. 2). The microprocessor chip optically communicates to the 1 MB memory array located remotely on a second identical chip an arbitrary distance away. The microprocessor sends requests (a ‘read’ or ‘write’), the memory address (location in memory to read or write), and write data (for write requests)

via the microprocessor-to-memory (P → M) link. The memory-to-microprocessor (M → P) link returns read data for read requests. A field programmable gate array (FPGA) provides the peripheral functionality of a motherboard, completing a user controllable computer.

For both P → M and M → P links, the laser light first couples into an electro-optic transmitter; laser light arriving in a single-mode fibre couples into an on-chip waveguide through a vertical grating coupler (VGC). The optical modulator, driven by circuits, modulates light in the waveguide and imprints it with on-off keyed binary data from the source. The light then exits the chip through a second vertical grating into a single-mode fibre bound for the other chip. Once there, the light couples into the receive site through a VGC, illuminates a receive photodetector, and is resolved back by the receiver circuit into binary data for the destination. The communication between the microprocessor and memory is full-duplex. Both P → M and M → P links run at 2.5 Gb s⁻¹, providing an aggregate 5 Gb s⁻¹ of memory bandwidth. Our demonstration uses only one wavelength of light; each additional wavelength increases the memory bandwidth by 5 Gb s⁻¹ for a total potential aggregate bandwidth of 55 Gb s⁻¹ without the need to use additional fibres.

A single 1,183-nm continuous-wave off-chip solid-state laser acts as the light source, with output power split 50/50 to share it across both the P → M and M → P links. To overcome the 4–6 dB coupling losses through each VGC due to unoptimized grating couplers, we insert an optical amplifier, which provides about 9 dB of gain, to obtain sufficient optical power at the receiver to resolve the signal. Using the optimized VGCs with losses of 1.2 dB (ref. 27) that exist as standalone test devices elsewhere on the same chip would eliminate the need for optical amplifiers in future design iterations.

To verify functionality of the photonically connected memory in the computer, we ran a combination of terminal-based and graphical programs (see Fig. 3b for an excerpt). To run a program, the control FPGA

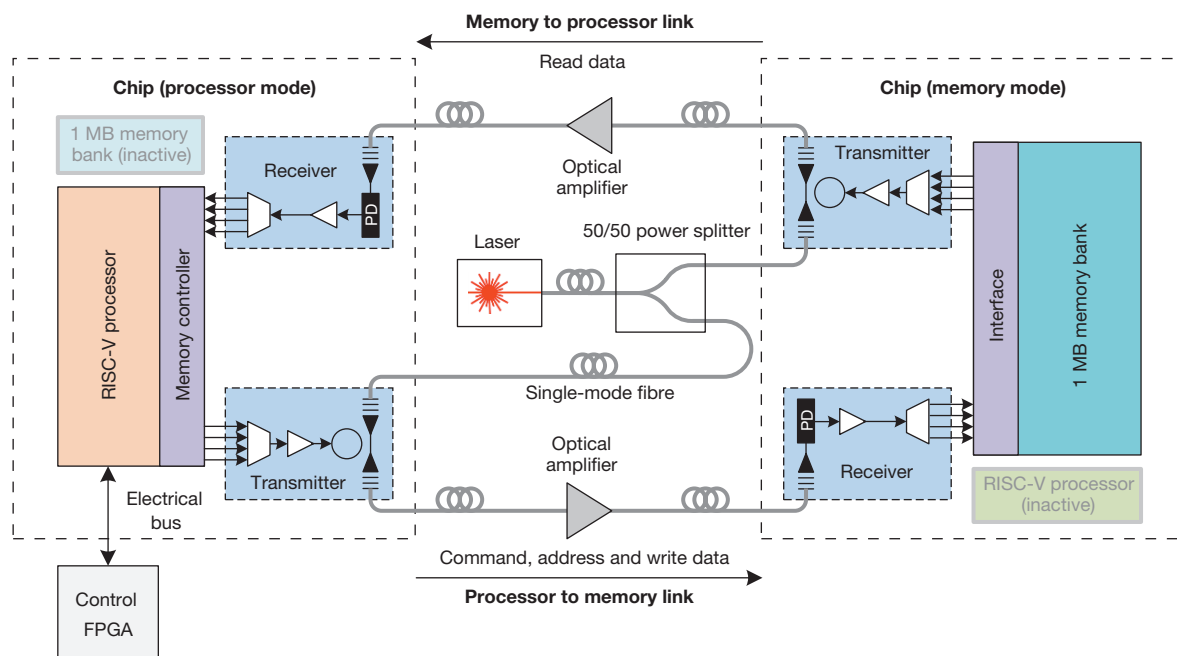


Figure 2 | Block diagram of the optical memory system. The system uses one chip acting as the processor and the other acting as memory, connected by a full-duplex optical link with a round-trip distance of 20 m by fibre. PD, photodetector.

first performs direct memory access through the memory controller to write all of the program's instructions into memory. Once the program is fully loaded, the FPGA issues a 'reset' signal to the processor and the processor begins execution of the program by fetching the first program instruction from memory (from address 0x00002000). During program execution, the processor writes and reads program data to

and from memory, in addition to reading the instructions from the memory. The control FPGA handles the printing of terminal outputs and acts as a display driver that reads from the frame buffer residing in memory to display a screen to the user. In all cases, the $P \rightarrow M$ and $M \rightarrow P$ optical links handle all communications to and from memory (which holds all the program instructions and data).

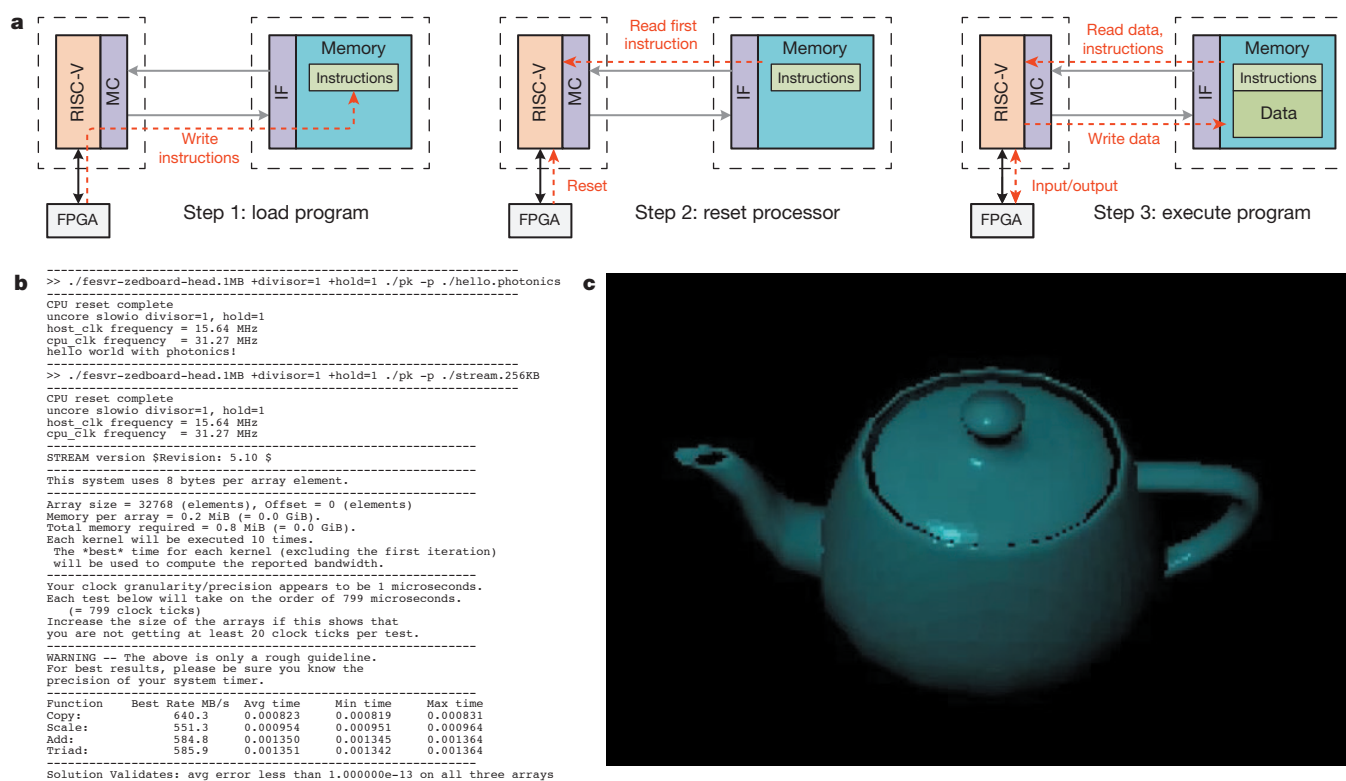


Figure 3 | Processor optical demonstration. **a**, Program loading and execution. MC is the memory controller in the processor; IF is the memory interface of the memory bank. **b**, Successful execution of the 'Hello world!' basic functionality test and the STREAM²⁸ memory

benchmark, two examples of terminal-based programs. **c**, Screen capture of the output of a three-dimensional teapot-rendering application running on the processor.

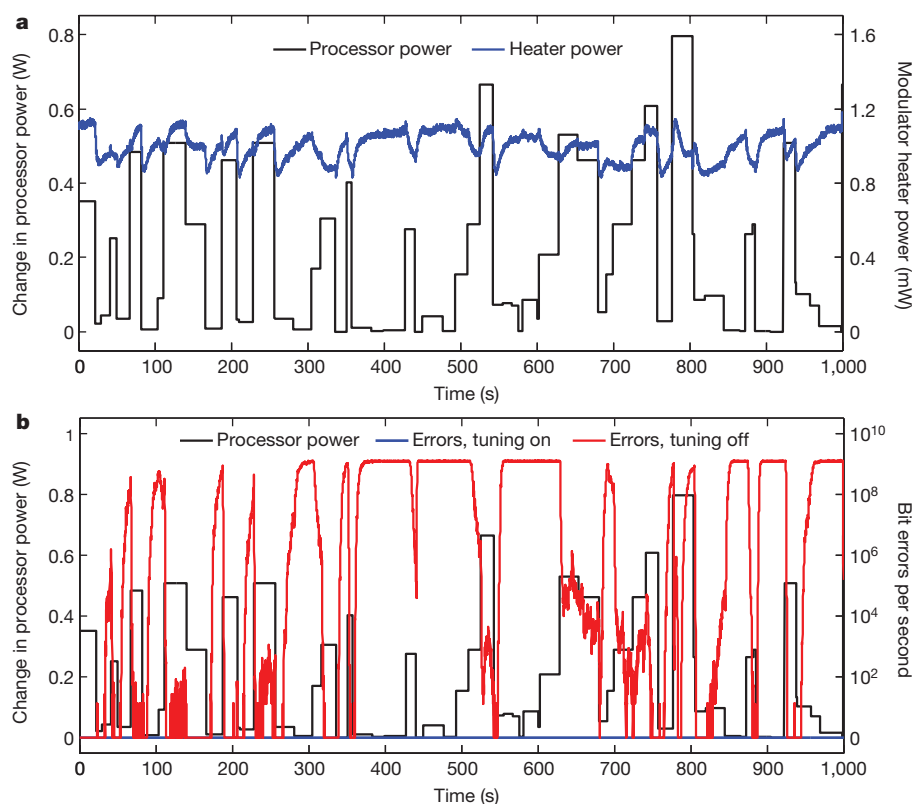


Figure 4 | Thermal-tuning stress test of the P → M link. a, Modulator heater output power with tuning switched 'on', overlaid on the power trace for the processor. The thermal-tuning controller changes the heater power output to adapt to the changes in temperature created by the changes in

processor power. **b,** Measured bit errors per second versus time with the thermal-tuning controller switched 'on' and 'off', overlaid on the power trace for the processor. The link with the tuning controller 'on' has no bit errors over the entire interval (a total of 2.5 Tb transmitted and received).

The processor clock frequency is locked to 1/80th of the aggregate bit rate of the P → M link (corresponding to a clock frequency of 31.25 MHz at 2.5 Gb s⁻¹) when demonstrating the processor using the optical link, the result of a decision that simplified engineering efforts during chip design. When operating in non-optical mode—by electrically communicating to the 1 MB bank of memory local to the same chip, or memory connected to the control FPGA by time-multiplexing memory data over the control interface—the processor can run at a maximum speed of 1.65 GHz. A demonstration of the system running these programs is provided in Supplementary Video 1.

To evaluate the robustness of the optical links and ring tuning control against thermal perturbations, we create a synthetic processor power trace by changing the voltage and frequency operating points of the processor (Fig. 4) over a 1,000-s period. The changes in processor power are representative of the behaviour of a processor as it runs different loads, affecting the chip temperature. The difference in temperature between the highest and lowest temperatures (processor at maximum and minimum power, respectively) is approximately 8 K. The thermal-tuning circuitry controls the output of the microheater integrated with the ring modulator to keep the resonant device locked to the laser wavelength, which keeps the link free of bit errors despite changes in temperature produced by the processor. With the tuning circuitry disabled, the same link experiences a number of bit errors depending on the processor power draw. The effect of thermal perturbations on the system during the execution of a program is shown in Supplementary Video 1.

Our demonstration of an electronic–photonic microprocessor chip could enable advances in very-large-scale integrated circuit (VLSI) technology, by adding nanophotonics as a new design dimension. Tailoring photonic devices to be integrated directly with electronics in an advanced-node CMOS process enabled a fully functioning electronic–photonic system on a single chip to be produced in a high-volume electronics foundry. The level of integration

allowed on-chip thermal-tuning control systems to guarantee robust operation of compact and energy-efficient, but also thermally sensitive, optical resonator devices, addressing one of the key remaining challenges for nanophotonic circuits adoption in VLSI technology.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 August; accepted 10 November 2015.

- Goodman, J. W., Leonberger, F. J., Kung, S.-Y. & Athale, R. A. Optical interconnections for VLSI systems. *Proc. IEEE* **72**, 850–866 (1984).
- Miller, D. A. Rationale and challenges for optical interconnects to electronic chips. *Proc. IEEE* **88**, 728–749 (2000).
- Young, I. *et al.* Optical I/O technology for tera-scale computing. *IEEE J. Solid-State Circ.* **45**, 235–248 (2010).
- Vantrease, D. *et al.* Corona: system implications of emerging nanophotonic technology. In *Proc. 35th Annual International Symposium on Computer Architecture (ISCA '08)* 153–164, <http://dx.doi.org/10.1109/ISCA.2008.35> (IEEE Computer Society, 2008).
- Shacham, A., Bergman, K. & Carloni, L. P. Photonic networks-on-chip for future generations of chip multiprocessors. *IEEE Trans. Comput.* **57**, 1246–1260 (2008).
- Batten, C. *et al.* Building manycore processor-to-DRAM networks with monolithic CMOS silicon photonics. *IEEE Micro* **29**(4), 8–21 (2009).
- Beamer, S. *et al.* Re-architecting DRAM memory systems with monolithically integrated silicon photonics. In *Proc. 37th Annual International Symposium on Computer Architecture (ISCA '10)* 129–140, <http://doi.acm.org/10.1145/1815961.1815978> (ACM, 2010).
- Xu, Q., Schmidt, B., Pradhan, S. & Lipson, M. Micrometre-scale silicon electro-optic modulator. *Nature* **435**, 325–327 (2005).
- Narasimha, A. *et al.* A 40-Gb/s QSPF optoelectronic transceiver in a 0.13 μm CMOS silicon-on-insulator technology. In *Optical Fiber Communication Conference OMK7*, <http://www.osapublishing.org/abstract.cfm?URI=OFC-2008-OMK7> (Optical Society of America, 2008).
- Assefa, S. *et al.* CMOS integrated nanophotonics: enabling technology for exascale computing systems. In *Optical Fiber Communication Conference OMM6*, <https://www.osapublishing.org/abstract.cfm?uri=OFC-2011-OMM6> (Optical Society of America, 2011).

11. Buckwalter, J., Zheng, X., Li, G., Raj, K. & Krishnamoorthy, A. A monolithic 25-Gb/s transceiver with photonic ring modulators and Ge detectors in a 130-nm CMOS SOI process. *IEEE J. Solid-State Circ.* **47**, 1309–1322 (2012).
12. Dupuis, N. *et al.* 30Gbps optical link utilizing heterogeneously integrated III-V/Si photonics and CMOS circuits. In *Optical Fiber Communications Conference, Th5A.6*, <http://dx.doi.org/10.1364/OFC.2014.Th5A.6> (Optical Society of America, 2014).
13. Orcutt, J. S. *et al.* Open foundry platform for high-performance electronic-photonic integration. *Opt. Express* **20**, 12222–12232 (2012).
14. Takahashi, O. *et al.* Migration of Cell broadband engine from 65nm SOI to 45nm SOI. In *International Solid-State Circuits Conference (ISSCC 2008) Dig. Tech. Pap.* 86–597, <http://dx.doi.org/10.1109/ISSCC.2008.4523069> (IEEE, 2008).
15. IBM Blue Gene team. Design of the IBM Blue Gene/Q Compute chip. *IBM J. Res. Dev.* **57**, 1–13 (2013).
16. Wendel, D. *et al.* The implementation of POWER7™: a highly parallel and scalable multi-core high-end server processor. In *International Solid-State Circuits Conference (ISSCC 2010) Dig. Tech. Pap.* 102–103, <http://dx.doi.org/10.1109/ISSCC.2010.5434074> (IEEE, 2010).
17. Waterman, A., Lee, Y., Patterson, D. A. & Asanović, K. *The RISC-V Instruction Set Manual, Volume I: User-Level ISA, Version 2.0*. Technical Report No. UCB/EECS-2014-54, <http://www.eecs.berkeley.edu/Pubs/TechRpts/2014/EECS-2014-54.html> (EECS Dept., Univ. California, Berkeley, 2014).
18. Narasimha, S. *et al.* High performance 45-nm SOI technology with enhanced strain, porous low-k BEOL, and immersion lithography. In *International Electron Devices Meeting (IEDM '06)* 1–4, <http://dx.doi.org/10.1109/IEDM.2006.346879> (IEEE, 2006).
19. Alloatti, L., Srinivasan, S., Orcutt, J. & Ram, R. Waveguide-coupled detector in zero-change complementary metal–oxide–semiconductor. *Appl. Phys. Lett.* **107**, 041104 (2015).
20. Sun, C. *et al.* A 45nm SOI monolithic photonics chip-to-chip link with bit-statistics-based resonant microring thermal tuning. In *2015 Symposium on VLSI Circuits C122–C123*, <http://dx.doi.org/10.1109/VLSIC.2015.7231348> (IEEE, 2015).
21. Shainline, J. M. *et al.* Depletion-mode carrier-plasma optical modulator in zero-change advanced CMOS. *Opt. Lett.* **38**, 2657–2659 (2013).
22. Wade, M. T. *et al.* Energy-efficient active photonics in a zero-change, state-of-the-art CMOS process. In *Optical Fiber Communication Conference Tu2E.7*, <http://dx.doi.org/10.1364/OFC.2014.Tu2E.7> (Optical Society of America, 2014).
23. Soref, R. A. & Bennett, B. Electrooptical effects in silicon. *IEEE J. Quantum Elect.* **23**, 123–129 (1987).
24. Selvaraja, S. K., Bogaerts, W., Dumon, P., Van Thourhout, D. & Baets, R. Subnanometer linewidth uniformity in silicon nanophotonic waveguide devices using CMOS fabrication technology. *IEEE J. Sel. Top. Quantum Elect.* **16**, 316–324, (2010).
25. Padmaraju, K., Chan, J., Chen, L., Lipson, M. & Bergman, K. Thermal stabilization of a microring modulator using feedback control. *Opt. Express* **20**, 27999–28008 (2012).
26. Sun, C. *et al.* A monolithically-integrated chip-to-chip optical link in bulk CMOS. *IEEE J. Solid-State Circ.* **50**, 828–844 (2015).
27. Wade, M. *et al.* 75% efficient wide bandwidth grating couplers in a 45 nm microelectronics CMOS process. In *2015 Optical Interconnects Conference 46–47*, <http://dx.doi.org/10.1109/OIC.2015.7115679> (IEEE, 2015).
28. McCalpin, J. D. STREAM: sustainable memory bandwidth in high performance computers. <http://www.cs.virginia.edu/stream/> (1995).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Twigg, Q. Nguyen, and M. Moreto Planas for help with processor infrastructure, A. Srinivasan for help with photodetector characterization, and S. Han for help with chip photos. This work was supported by DARPA POEM award HR0011-11-C-0100, led by J. Shah and DARPA PERFECT award HR0011-12-2-0016, led by J. Cross. We thank M. Casper, J. Torneden, and the team at the Kansas City Plant for their support of our design submissions over the years leading up to this work. Support is also acknowledged from the Berkeley Wireless Research Center, UC Berkeley ASPIRE Lab, MIT CICS, National Science Foundation, FCRP IFC, Trusted Foundry, Intel, Santec, and NSERC. The views expressed are those of the authors and do not reflect the official policy or position of the DoD or the US Government.

Author Contributions C.S. developed the thermal tuning circuitry, designed the memory bank, implemented the ‘glue-logic’ between various electronic components, and performed top-level assembly of electronics and photonics. M.T.W. optimized modulator designs for thermal tuning, designed the grating couplers, and performed top-level assembly of photonics regions used in our demonstration. C.S. and Y.L. designed the system-level architecture and demonstrated the processor with photonic input/output. Y.L. wrote and/or adapted the test programs for the processor demonstration. Y.L. and A.S.W. developed the RISC-V ISA and processor implementation. J.S.O. created the CAD infrastructure for photonic layouts, designed the photodetector used in our demonstration, and assembled initial photonic layouts and passive devices. L.A. improved the CAD infrastructure, developed new rules for design rule checking, and contributed new photodetector designs. C.S., M.T.W., Y.L., and L.A. contributed to chip verification and testing. M.S.G. designed and implemented the receiver circuit. J.M.S. designed, implemented, and tested the original version of the modulator. R.R.A. performed the physical implementation of the processor and designed the chip and adapter printed circuit boards. S.L. developed the selective substrate removal process and contributed to the thermal tuning method. B.R.M. assisted with chip implementation and performed initial substrate removal experiments. R.K. assisted in the rework of new grating coupler designs. F.P. contributed to layout and analysis for couplers and modulators. A.H.A. created new photodetector designs. H.M.C. and A.J.O. assisted with processor design. J.C.L. and Y.-H.C. contributed components in the transceiver regions. V.M.S., M.A.P., R.J.R., and K.A. supervised the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to V.M.S. (vlada@berkeley.edu), M.A.P. (milos.popovic@colorado.edu), R.J.R. (rajeev@mit.edu) or K.A. (krste@berkeley.edu).

METHODS

Chip implementation. The key chip characteristics are summarized in Extended Data Table 1. Photonic devices were prepared in Cadence Virtuoso (an industry-standard design tool for frontend electronics) in conjunction with mixed-signal electronics²⁹. Digital electronics were implemented using a combination of digital-synthesis and place-and-route tools from Synopsys and Cadence. All photonic and electronic designs conform to the CMOS manufacturing rules (more than 5,000 rules) of IBM's commercial 45-nm thin buried-oxide SOI process (12SOI), with physical verification performed using Mentor Graphics Calibre.

Chip fabrication. The chips were fabricated through the standard 12SOI process flow. We submitted our design for mask aggregation through the Trusted Access Program Office (TAPO) shuttle run, with the chip mask set treated as if it were an ordinary electronics design. The physical design dimensions, including the cross-sectional layer type and thickness information not reported here, are provided as part of the standard electronic design kit that is made available to IBM foundry customers under a non-disclosure agreement. A subset of process and performance information regarding this process can be found in various official IBM publications on electronic CMOS process development^{18,30,31}.

Electrical packaging. The chips from the foundry are bumped with controlled-collapse chip connection (C4) solder balls. The chips are then flip-chip mounted (the chip's substrate is exposed on top) to an 8-layer FR4 printed circuit board through C4 solder reflow. This forms all 249 electrical connections (including power and ground) from the chip to the printed circuit board. Epoxy encapsulation is added to the mounted chips for additional mechanical support and to protect the mounted chips. These steps are typical for an electrical chip package and were performed by CVInc.

Patterned substrate removal of a packaged chip. The electrically packaged samples are first backside-ground to thin the chip substrate down to 100–150 μm (performed by Aptek Industries). We then clean the backside surface with isopropyl alcohol and an N_2 air gun. We next apply Kapton tape over the substrate regions that we do not wish to remove (over the processor and the 1 MB memory bank). Afterwards, the chips are placed in a chamber that supplies XeF_2 gas to isotropically etch the silicon substrate, removing it as the volatile product SiF_4 . We use a pulsed-etch technique, in which etch steps of 120 s were interleaved with 60-s periods during which we pump out the reaction products. The pressure used in the chamber is 3.4 Torr. Because electronics are unaffected by the substrate removal, the very coarse feature definition provided by tape and hand alignment is sufficient. On average, the substrate removal process takes 10–30 cycles (depending on the thickness after the backside grind) with a success yield of 80% (defined as having a working processor after substrate removal). We stop the etch when the substrate over the desired etch region has disappeared when inspected by eye. The steps above are easily implementable at wafer scale in high-volume manufacturing using standard photolithographic techniques³², which can also improve uniformity and yield of the post-processing as well as the resolution and alignment of the etch regions.

Optical testing. The 1,183-nm laser is a quantum dot DFB (distributed feedback) laser available from QDLaser. We used lensed fibres available from Oz Optics with a spot size of 5 μm and a working distance of 26 μm to couple light into the VGCs through the chip backside (after substrate removal). The spot size is matched to the 5- μm mode-field diameter of the VGCs. We used 3-axis positioner stages (Thorlabs NanoMax) to position and align fibres over the grating couplers of the test sites. The shown demonstrations require a total of 3 fibres coupled to each chip. Minimum fibre-to-coupler insertion loss was achieved by angling the fibres at 19° off-normal from the surface of the chip. To adjust the polarization of the input light, we use 3-paddle manual polarization controllers from Thorlabs (although these can be avoided if using polarization-maintaining fibres). For this first demonstration, we chose the manual fibre alignment approach to freely couple into any of the hundreds of optical test sites located throughout the chip. To make a permanent fibre-attach, we could leverage commercial optical packaging techniques for VGCs, such as through horizontal fibre array blocks with angle-cleaved fibres³³ or through vertical fibre array pigtailed^{9,34}.

Processor testing. The control FPGA is a Zedboard FPGA, providing an intermediate hardware interface between the electrical links of the processor and an ethernet connection to the laboratory control computer. The individual cores incorporate a 64-bit scalar core, floating-point unit, vector accelerator, and private caches³⁵. Programs are compiled from C source code using a GCC-based C compiler targeted for the RISC-V ISA. The implementations of the RISC-V processor and the software compilation stack are available at <http://bar.eecs.berkeley.edu/projects.html>. Details of the RISC-V ISA standard are found at <http://www.riscv.org>. The full system is stable and can execute an arbitrary

number of programs. A representative set of programs tested on this processor is as follows.

- 'Memory test'. The control FPGA writes to and reads from every location in memory through direct memory access to verify that the memory interface is fully functional and that all bits are correct. The processor is idled for this test.
- 'Hello world!'. A program that asks the processor to print out a single line of text to the terminal, which is sent to the control FPGA to be displayed to the user.
- 'STREAM'. A popular memory benchmarking application²⁸; the outputs of the program are printed to the terminal and displayed to the user.
- 'Teapot renderer'. A program that pixel shades a three-dimensional teapot using the Blinn–Phong shading model and outputs the rendered image. The location and colour of the light source illuminating the teapot in the rendered image is controlled by the user using the keyboard. The processor performs all calculations and writes the image to the frame buffer in memory using the optical links. It then reads the content of the frame buffer over the optical link and sends it to the control FPGA to display it as an image to the user.
- 'Linux'. A full Linux operating system. Once Linux boots, the user is free to run any program, including 'python', 'top', or file system operations (the file system behaviour is coordinated by the control FPGA). This test uses memory connected to the control FPGA and not the optically connected memory, because the memory footprint of the Linux kernel is too big to fit in the 1 MB memory bank.

The 1:80 ratio between processor clock frequency and the $\text{P} \rightarrow \text{M}$ link throughput was chosen to keep processor frequency reasonable if the links operated at higher data rates than anticipated at design time and when all wavelengths in the $\text{P} \rightarrow \text{M}$ link are active. For example, if the $\text{P} \rightarrow \text{M}$ link supported an 80 Gb s^{-1} aggregate data rate, then the processor needs to operate at 1.0 GHz, which is well within its abilities. Alternatively, if the ratio was 1:10, then the processor would need to operate at 8 GHz, which is impractical.

Transmitter and receiver circuit specifications. At the 2.5 Gb s^{-1} operating point used in our demonstration, the transmitter uses the 1-V digital supply, which corresponds to a transmitter energy of 20 fJ per bit and achieves an insertion loss of 3 dB at an on:off ratio of 6 dB for non-return-to-zero binary data. The modulator is effectively 'driverless' insofar as no analogue driver electronics are needed to bridge between digital logic and the optical modulator, owing to the efficiency of the latter. The thermal tuning for the modulator ring consumes a fixed 192 fJ per bit for the control circuit and 0–2.5 mW for the heater power, dependent on the tuned range (for the heater output power of 1.5 mW in Supplementary Video 1, this corresponds to 600 fJ per bit). More detailed transmitter and thermal tuner descriptions have been reported previously²⁰. The receiver has a 10^{-12} bit-error-rate sensitivity (OMA) of -5 dBm up to 5 Gb s^{-1} , degrading to -3.8 dBm at 8 Gb s^{-1} , and -0.8 dBm at 10 Gb s^{-1} . At 2.5 Gb s^{-1} , the receiver energy efficiency is 496 fJ per bit, improving to 297 fJ per bit at 10 Gb s^{-1} . Summing up, we report a total circuit energy efficiency of 1.3 pJ per bit at 2.5 Gb s^{-1} —a power consumption of 3.25 mW. The bandwidth density of the transceivers is approximately 300 $\text{Gb s}^{-1} \text{ mm}^{-2}$ of chip area. The key specifications are summarized in Extended Data Table 2.

Link specifications. In the 2.5 $\text{Gb s}^{-1} \text{ P} \rightarrow \text{M}$ and $\text{M} \rightarrow \text{P}$ links used in our demonstration, the transmitter input VGC, transmitter output VGC, and receiver input VGC contribute 4 dB, 4 dB, and 6 dB of link insertion loss, respectively. The 1,183-nm laser outputs 9.2 dBm such that 5.2 dBm (50/50 split, with an approximately 1-dB excess loss of the splitter) is incident upon each of the input transmit VGC of each link ($\text{P} \rightarrow \text{M}$ and $\text{M} \rightarrow \text{P}$). At this laser power level, the OMA of each transmitter is -7 dBm , with an average optical power of -9 dBm . Each amplifier adds 9 dB of optical gain, completing the $\text{P} \rightarrow \text{M}$ and $\text{M} \rightarrow \text{P}$ links each with an extra 1-dB link margin. A chip iteration incorporating 1.2-dB-loss VGCs²⁷ into the $\text{P} \rightarrow \text{M}$ and $\text{M} \rightarrow \text{P}$ link would remove 10.4 dB of excess insertion loss. These devices were high-risk test structures on the current chip and so were not placed in the $\text{P} \rightarrow \text{M}$ and $\text{M} \rightarrow \text{P}$ transceivers. Using these couplers at the same input laser power level as before, both links could complete, without an amplifier, with an extra 2.4 dB of link margin. The 1,183-nm laser was made by QDLaser and uses 55 mA of pump current at a laser diode bias of 1.3 V to output 9.2 dBm (8.3 mW) of power. This corresponds to a power use of 71.5 mW and a wall-plug efficiency of 11.6%. The laser is shared across both $\text{P} \rightarrow \text{M}$ and $\text{M} \rightarrow \text{P}$ links and the total wall-plug energy efficiency (laser and circuit) is 15.6 pJ per bit. The laser has a threshold of 29 mA and a slope efficiency of 0.32 mW mA^{-1} .

Potential for improved performance. The chip demonstrated here is a first working research prototype, and the current achieved performance is by no means representative of the absolute performance limits of this technology. We describe a few known ways to improve performance in the following.

(1) The current modulator design uses a mid-level p-implant (10^{17} – 10^{18} cm^{-3}) for p-contacts as opposed to a p+ implant, creating high series contact resistance

that limits its bandwidth. Future design iterations will use the p+ implant to improve device bandwidth. Moreover, the modulators used only two out of several different doping implants available in the process for different transistors and transistor thresholds. Substantial improvements may be possible with other available implants.

(2) The current detector is absorption-length limited¹⁹ and resonating the detector can improve sensitivity without an increase in the device size. Resonant detectors, implemented as a spoked-ring cavity in a manner similar to that of the modulator, exist on the same chip as standalone devices in the independent-device and transceiver regions. If incorporated with processor and memory transceivers in a future chip, they would improve sensitivity by approximately 6 dB (to an OMA sensitivity of -11 dBm), which would be competitive with state-of-the-art integrated receivers. In addition, the current receiver circuit design is very conservative and could be optimized to further improve the sensitivity by 6 dB. The circuit could also be placed closer to the photo-detector to minimize wiring capacitance.

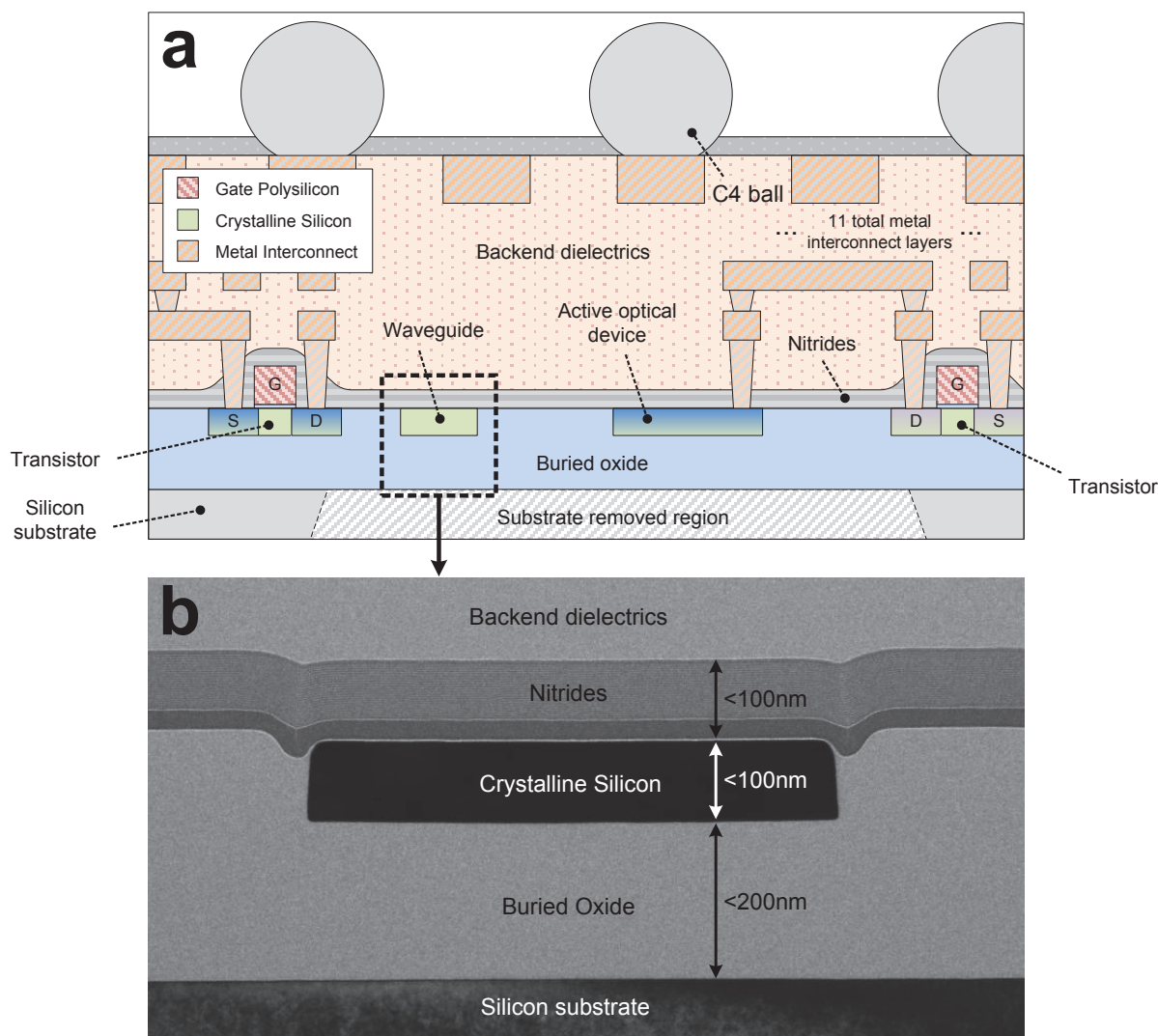
(3) Our demonstration uses the laser at a power level far below that for peak efficiency, which is 16% at 30 mW. Operation of the current laser at the peak-efficiency power level and sharing of the output power across multiple links on the chip or, alternatively, usage of a laser optimized for the given output power are techniques for improving the energy efficiency of the link, even without any device improvements.

Applicability to CMOS processes with bulk silicon substrates. CMOS processes using a bulk silicon substrate lack a patternable crystalline silicon layer, which motivates the use of alternative devices in polycrystalline silicon and a small number of process changes³⁶. However, some guiding principles of zero-change integration, such as reuse of existing transistor mask levels, repurposing of transistor materials for optics, and compact integration using silicon micro-rings, can be

applied to minimize changes to the process frontend, which are the most harmful to process-native electronics. These concepts have been applied successfully in practice to enable functional photonics in bulk^{26,36}, although at a far smaller scale than is demonstrated here.

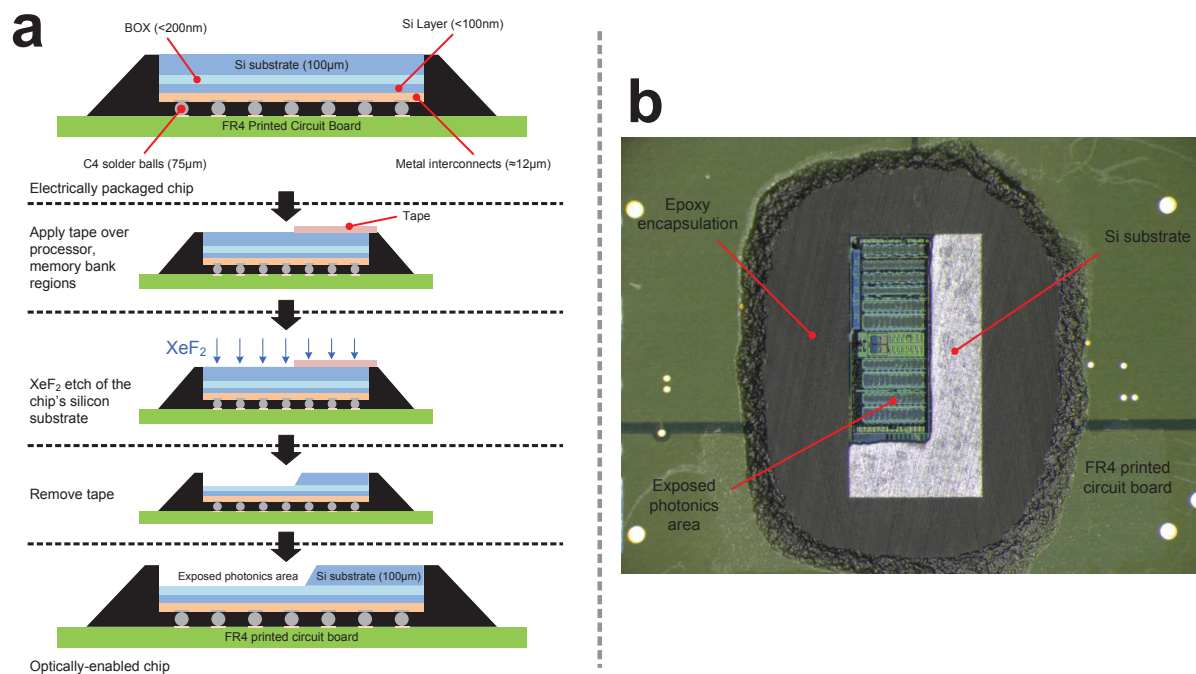
Code availability. The source code for the processor is available at <http://bar.eecs.berkeley.edu/projects.html>. Other test applications (such as STREAM) can be found at the respective cited references.

29. Orcutt, J. S. & Ram, R. J. Photonic device layout within the foundry CMOS design environment. *IEEE Photonic. Tech. Lett.* **22**, 544–546 (2010).
30. Kalla, R., Sinharoy, B., Starke, W. J. & Floyd, M. Power7: IBM's next-generation server processor. *IEEE Micro* **30**, 7–15 (2010).
31. Lee, S. *et al.* Record RF performance of 45-nm SOI CMOS technology. In *International Electron Devices Meeting (IEDM 2007)* 255–258, <http://dx.doi.org/10.1109/IEDM.2007.4418916> (IEEE, 2007).
32. Roger, A. Breaking a new sound barrier: it's a mic-on-a-chip. *Electron. Des.* **54**, <http://electronicdesign.com/analog/breaking-new-sound-barrier-its-mic-chip> (2006).
33. Pavarelli, N., Lee, J. S. & O'Brien, P. A. Packaging challenges for integrated silicon photonic circuits. In *Proc. SPIE Vol. 9133* (eds Vivien, L. *et al.*) 91330F, <http://dx.doi.org/10.1117/12.2058559> (SPIE, 2014).
34. Kopp, C. *et al.* Silicon photonic circuits: on-CMOS integration, fiber optical coupling, and packaging. *IEEE J. Sel. Top. Quant. Elect.* **17**, 498–509 (2011).
35. Lee, Y. *et al.* A 45nm 1.3GHz 16.7 double-precision GFLOPS/W RISC-V processor with vector accelerators. In *40th European Solid State Circuits Conference (ESSCIRC 2014)* 199–202, <http://dx.doi.org/10.1109/ESSCIRC.2014.6942056> (IEEE, 2014).
36. Meade, R. *et al.* Integration of silicon photonics in bulk CMOS. In *2014 Symposium on VLSI Technology Dig. Tech. Pap.* 228–229, <http://dx.doi.org/10.1109/VLSIT.2014.6894427> (IEEE, 2014).



Extended Data Figure 1 | Chip cross-section. a, Full chip cross-section (not to scale) from the silicon substrate to the C4 solder balls, showing the structures of electrical transistors, waveguides, and contacted optical devices. G, S, and D mark the structures that form the gate, source, and drain, respectively, of an electrical transistor. The minimum separation

between transistors and waveguides is $<1\ \mu\text{m}$, which is set only by the distance at which evanescent light from the waveguide begins to interact with the structures of the transistor. **b**, Transmission electron microscopy cross-section micrograph of an optical waveguide, before substrate removal.



Extended Data Figure 2 | Selective substrate removal. **a**, Selective substrate removal steps for the flip-chip packaged chip, using tape as a coarse mask for defining areas that retain the substrate. BOX, buried oxide. **b**, Photo of a selective-substrate-removed fully electrically packaged electronic-photonic processor chip.

Extended Data Table 1 | Summary of chip characteristics

Characteristic	Value
Number of Transistors	70 Million
in Processor/Memory	60 Million
in P2M/M2P Transceivers	≈4 Million
in Standalone Transceivers	≈5 Million
Number of VGCs, Rings, PDs	851
in P2M/M2P Transceivers	324
in Standalone Sites	527
Processor Cores	2
Max Processor Frequency	1.65 GHz
L1 Instruction Cache	2 × 16 KB
L1 Data Cache	2 × 32 KB
L1 Vector Instruction Cache	2 × 8 KB
Memory Bank	1 MB
Number of P2M/M2P Transceiver Banks	4
Transmitters/Receivers per Bank	11
Max Wavelengths	11
Theoretical throughput if all transceivers on the chip were active	550 Gb/s Tx 900 Gb/s Rx

Summary of the physical characteristics of the chip, such as the total number of electrical and optical devices, the processor parameters, and the configurations of the photonic transceiver banks connected to the processor and memory. L1 is the level 1 processor cache, Tx refers to transmit, and Rx refers to receive.

Extended Data Table 2 | Summary of transceiver performance

Property	P2M/M2P Transceivers	Standalone Transceivers
Waveguide Loss	4.3 dB/cm	4.3 dB/cm
Grating Coupler Loss	4 dB and 6 dB	1.2 dB
Tx Data Rate	2.5 Gb/s	5 Gb/s
Tx Extinction Ratio	6 dB	>6 dB
Tx Insertion Loss	3 dB	3 dB
Tx Power	0.02 pJ/bit	0.03 pJ/bit
Rx Data Rate	2.5 Gb/s	10 Gb/s
PD Responsivity	0.023 A/W	0.10 A/W
Rx OMA Sensitivity	−5 dBm@2.5 Gb/s	−7.2 dBm@10 Gb/s*
Rx Power	0.50 pJ/bit	0.30 pJ/bit
Ring Tuning Range	≈3.0 nm	≈3.0 nm
Ring Heater Tuning Efficiency	1.25 nm/mW	1.25 nm/mW
Ring Tuning Control Power	0.19 pJ/bit	0.14 pJ/bit

* Estimated OMA sensitivity using the 0.10 A/W PD.

Summary of the performance metrics of the P → M and M → P transceivers, and of the transceivers that exist in the standalone independent devices region. PD, photodetector.

Processing and properties of magnesium containing a dense uniform dispersion of nanoparticles

Lian-Yi Chen^{1,2,3}, Jia-Quan Xu², Hongseok Choi⁴, Marta Pozuelo², Xiaolong Ma⁵, Sanjit Bhowmick⁶, Jenn-Ming Yang², Suveen Mathaudhu⁷ & Xiao-Chun Li^{1,2}

Magnesium is a light metal, with a density two-thirds that of aluminium, is abundant on Earth and is biocompatible; it thus has the potential to improve energy efficiency and system performance in aerospace, automobile, defence, mobile electronics and biomedical applications^{1–5}. However, conventional synthesis and processing methods (alloying and thermomechanical processing) have reached certain limits in further improving the properties of magnesium and other metals⁶. Ceramic particles have been introduced into metal matrices to improve the strength of the metals⁷, but unfortunately, ceramic microparticles severely degrade the plasticity and machinability of metals⁷, and nanoparticles, although they have the potential to improve strength while maintaining or even improving the plasticity of metals^{8,9}, are difficult to disperse uniformly in metal matrices^{10–14}. Here we show that a dense uniform dispersion of silicon carbide nanoparticles (14 per cent by volume) in magnesium can be achieved through a nanoparticle self-stabilization mechanism in molten metal. An enhancement of strength, stiffness, plasticity and high-temperature stability is simultaneously achieved, delivering a higher specific yield strength and higher specific modulus than almost all structural metals.

We describe a processing method of achieving a uniform dispersion of dense silicon carbide (SiC) nanoparticles in magnesium (Mg) matrix (see Methods and Extended Data Fig. 1). Ingots of Mg6Zn (1 vol% SiC) were first obtained. At this stage, the nanoparticles were mostly distributed along the grain boundary region owing to the pushing of nanoparticles by the solidification front¹⁵. Previous work showed that a higher drag force in the melt via dense nanoparticles (for example, >6 vol%) could promote nanoparticle engulfment¹⁶. Typically, a low volume fraction of nanoparticles can be uniformly dispersed by ultrasonic processing, but this is not effective for dispersing dense nanoparticles¹⁶. Therefore, nanoparticles were concentrated by evaporating away magnesium and zinc from the Mg6Zn (1 vol% SiC) ingot at 6 torr in a vacuum furnace. After evaporation and slow cooling at approximately 0.23 K s^{-1} , a sample with about 14 vol% nanoparticles in an Mg2Zn matrix was obtained.

We first characterized the distribution and dispersion of SiC nanoparticles in as-solidified magnesium samples using scanning and transmission electron microscopy (SEM, TEM). To clearly reveal the nanoparticles, the SEM samples were cleaned by low-angle ion milling (10° , to remove the nanometre-sized polishing powders) and then slightly etched by gallium ions (90° , to preferentially etch the magnesium matrix) using a focused ion beam (FIB). SEM images in Fig. 1a and b were acquired at a 52° tilt to expose the nanoparticles on the surface of the magnesium matrix. This high volume fraction of nanoparticles is uniformly distributed and dispersed in the magnesium matrix, as shown in Fig. 1a and b. This uniform dispersion of nanoparticles in the magnesium matrix is also confirmed by TEM analysis

(Extended Data Fig. 2a). Some of the nanoparticles located at different depths in the thin film appear overlapped in this bright-field TEM image, owing to the conventional transmission mode in the TEM. Extended Data Fig. 2b shows a histogram indicating the SiC nanoparticle size distribution with an average diameter of 60 nm. Additionally, we performed energy-dispersive X-ray spectroscopy and microhardness tests on different parts of the sample. Extended Data Figure 2c and d shows that the silicon concentration (in weight per cent) and the microhardness value are both uniform from top to bottom and from the centre to the edge of the sample, which validates the uniform distribution of dense SiC nanoparticles in the magnesium matrix after solidification.

The uniform dispersion of nanoparticles in the as-solidified samples provides convincing evidence that SiC nanoparticles were previously well dispersed and self-stabilized in the molten magnesium before solidification. Prior studies reported that ceramic nanoparticles tend to form microclusters and then segregate after ultrasonic processing stops, mostly due to attractive van der Waals forces between nanoparticles¹⁷. Surprisingly, even though our samples remained in the liquid state without ultrasonic processing for about 4 h, the nanoparticles were still uniformly dispersed. Our theoretical analysis of the process physics, detailed in Methods subsection ‘Nanoparticle self-stabilization mechanism’, suggests that the self-stabilization of dense nanoparticles is attributed to three major factors (as schematically shown in Fig. 1c):

(1) A wetting angle of 83° between SiC nanoparticles and molten magnesium at the processing temperature, which creates an energy barrier of $3.87 \times 10^4\text{ zJ}$ to prevent an atomic-scale contact and sintering of SiC nanoparticles in the melt.

(2) A small attractive van der Waals potential, about -12.17 zJ at the secondary minimum in Fig. 1c, between SiC nanoparticles in magnesium melt, caused by a small difference in the Hamaker constants between SiC and molten magnesium.

(3) A high thermal energy of about 13.8 zJ , such that SiC nanoparticles can overcome the van der Waals attraction in magnesium melt.

Therefore, the repulsive energy barrier to prevent SiC nanoparticles from contact and sintering is much higher than the thermal energy, which is the main driving force that effectively separates SiC nanoparticles in the magnesium melt. While the attractive van der Waals potential tries to hold the SiC nanoparticles together into quasi-clusters, the high thermal energy allows the nanoparticles to break free from their attraction, resulting in dispersed nanoparticles in the melt. This thermally activated dispersion and self-stabilization mechanism provides a new pathway to achieve a uniform dispersion of dense nanoparticles in liquids when a repulsive force is not available through conventional methods.

In addition to a uniform dispersion of the nanoparticles, the interface between the matrix and the reinforcements plays a key part in

¹Scifaturing Laboratory, Department of Mechanical and Aerospace Engineering, University of California, Los Angeles, California 90095, USA. ²Department of Materials Science and Engineering, University of California, Los Angeles, California 90095, USA. ³Department of Mechanical and Aerospace Engineering, Missouri University of Science and Technology, Rolla, Missouri 65409, USA. ⁴Department of Mechanical Engineering, Clemson University, Clemson, South Carolina 29634, USA. ⁵Department of Materials Science and Engineering, North Carolina State University, North Carolina 27695, USA. ⁶Hysitron Inc., Minneapolis, Minnesota 55344, USA. ⁷Department of Mechanical Engineering, University of California, Riverside, California 92521, USA.

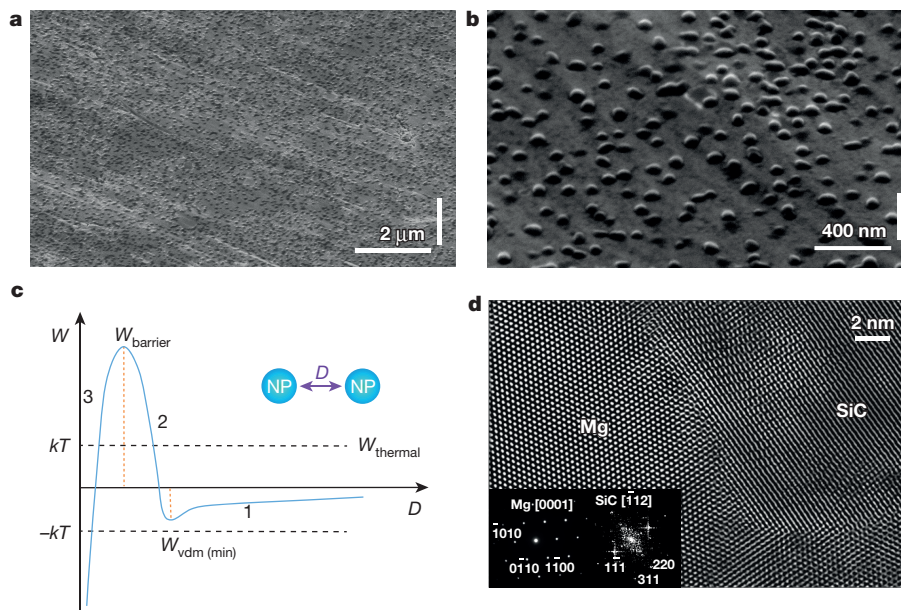


Figure 1 | Uniform dispersion of SiC nanoparticles in as-solidified magnesium alloy matrix. **a, b,** SEM images of the Mg₂Zn (14 vol% SiC) sample acquired at a 52 °C tilt angle and at different magnifications showing the uniform distribution and dispersion of SiC nanoparticles in the magnesium matrix. **c,** The principle of thermally activated dispersion and stabilization. The interaction potential W for two SiC nanoparticles (NPs; blue circles, separated by a distance D) that interact inside the magnesium melt is shown as the blue curve, which has three segments (labelled). Segment 1 is dominated by van der Waals interaction, segment 2 is dominated by the

interfacial energy increase when the Mg–SiC interface is replaced by SiC surfaces, and segment 3 is the interfacial energy drop due to SiC nanoparticles contacting and sintering. $W_{\text{vdw (min)}}$ is the minimum van der Waals potential for maximum attraction, W_{barrier} is the energy barrier due to the interfacial energy increase, $W_{\text{thermal}} = kT$ is the thermal energy. **d,** Fourier-filtered atomic-resolution TEM image showing a characteristic interface between a SiC nanoparticle and the magnesium matrix. Insets are the fast Fourier transforms of the magnesium matrix (left) and the SiC nanoparticle (right), oriented to the [0001] and [112] zone axes, respectively.

the development of high-performance nanocomposites. In this study, the interfaces between SiC nanoparticles and the magnesium matrix were characterized at the atomic scale by high-resolution TEM. Semi-coherent bonding between SiC nanoparticles and magnesium was observed (Fig. 1d), which should result in a strong interfacial bonding.

To determine the property enhancement induced by these dense dispersed nanoparticles, we first conducted *in situ* SEM micropillar compression tests at room temperature, as shown in Fig. 2. Single-crystal micropillars with diameters and lengths of 4 μm and 8 μm , respectively, were machined by FIB from the as-solidified samples, with average grain sizes of $1,011 \pm 265 \mu\text{m}$ and $23.6 \pm 14.1 \mu\text{m}$ for Mg₂Zn and Mg₂Zn (14 vol% SiC), respectively. The single-crystal nature of the micropillar is shown in Fig. 2d. This set of testing was designed only to evaluate the effect of nanoparticles on the strengthening without considering any influence of grain boundaries¹⁸. Micropillar size was also carefully selected to avoid size-induced strengthening, as we demonstrate later, and to provide results comparable with those from macroscale tests of magnesium alloys¹⁹. Additionally, these micropillars were machined with a particular orientation in order to induce deformation by basal slip and evaluate the effect of the nanoparticles on the weakest slip system in magnesium.

As an example, the micropillar in Fig. 2c is a single crystal oriented to the [2110] zone axis, as shown by the selected area electron diffraction pattern in Fig. 2d. Two rings corresponding to the SiC nanoparticles are also identified. In addition, it can be seen that the basal direction is forming an angle of 65° with the loading direction. Under this orientation, the deformation mechanism by basal slip will be favoured, as predicted by the Schmid factor²⁰.

The results from microcompression tests (Fig. 2a) show that the Mg₂Zn samples without nanoparticles yield at only around 50 MPa, then experience repeated loading–unloading cycles owing to severe basal slipping, as shown in Fig. 2 and Supplementary Video 1. In contrast, the samples with nanoparticles yield at a much higher strength of around 410 MPa, and bear a gradually increasing load smoothly to

a plastic strain of over 30%, as shown in Fig. 2a and Supplementary Video 2. Moreover, after deformation, multiple slip traces are observed in the samples without nanoparticles (Fig. 2b), but only one major slip trace that developed at the later stage of deformation is observed in the samples with nanoparticles (Fig. 2c). Even after the formation of the major slip trace at the last stage of deformation, the samples with nanoparticles can still bear load smoothly. These results demonstrate that the dense dispersed nanoparticles can not only greatly strengthen the material but can also enable a more uniform and stable deformation.

To study the deformation mechanisms of the micropillars under compression, TEM samples of the compressed micropillars were also prepared by FIB. Figure 2e and f shows the TEM analysis of the compressed micropillar with SiC nanoparticles shown in Fig. 2c. A Fourier-filtered high-resolution TEM image (Fig. 2e) reveals partial dislocations on the basal planes of the magnesium matrix, as confirmed by its indexed fast Fourier transform (Fig. 2f). In fact, basal dislocations are surrounding a SiC nanoparticle of around 40 nm size (inset to Fig. 2e) with their (111) planes parallel to the (0002) planes of the magnesium matrix. The TEM study suggests that the densely dispersed nanoparticles might effectively block the dislocation slip on the basal planes, which could explain the suppression of the characteristic multiple slip bands observed in the samples without nanoparticles (Fig. 2b). At the yield point, the sample tends to slip along the weakest planes, but the strong SiC nanoparticles might block further slip along those planes. Since the density of the nanoparticles is very high, slip along weaker atomic planes can be effectively blocked. This would prevent a localized deformation along weaker atomic planes and would enable the activation of slip along other atomic planes.

We note that twinning—another important deformation mechanism in magnesium—has not been found in the 4- μm Mg₂Zn (14 vol% SiC) micropillars after compression. Previous studies have shown that twinning in magnesium can be suppressed by reducing both sample and grain size (to around 2–3 μm)²¹ or by the presence of fine dispersed particles^{22,23}. To rule out the possibility of suppression of twinning by sample size, we conducted further experiments on larger micropillars of

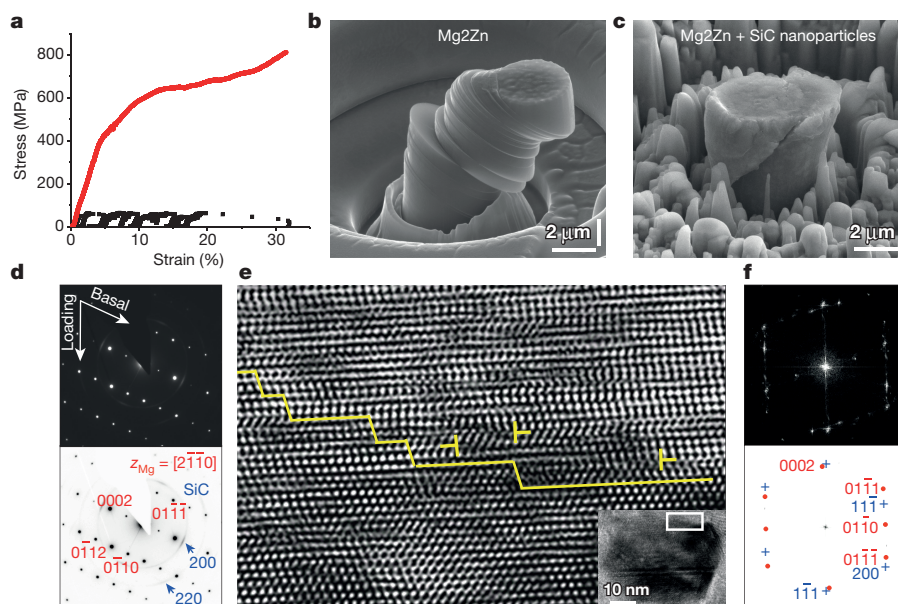


Figure 2 | Mechanical behaviour of as-solidified samples at room temperature. **a**, Engineering stress–strain curves of micropillar as-solidified samples without (black) and with (red) nanoparticles. **b**, **c**, SEM images showing the morphology of post-deformed samples without (**b**) and with (**c**) nanoparticles. **d**, The top panel shows a representative selected area electron diffraction pattern taken from a thin-film FIB prepared from the Mg₂Zn (14 vol% SiC) micropillar shown in **c**. Its colour-coded indexed selected area electron diffraction pattern (bottom panel) reveals a single-crystal micropillar oriented to the $[2\bar{1}\bar{1}0]$ zone axis of the magnesium matrix, z_{Mg} . Note that the basal direction is forming an angle of 65° with the loading direction. **e**, Fourier-filtered high-resolution TEM image of the region marked by a white rectangle in

the inset. The inset is a bright-field TEM image of a SiC nanoparticle. The SiC–Mg interface is highlighted by a solid yellow line. Partial dislocations (marked in yellow) terminating in stacking faults are located on the basal planes. Interface steps take place on the $\{111\}$ planes. **f**, The top panel shows the fast Fourier transform of the image in the inset to **e** and the bottom panel shows its colour-coded indexed fast Fourier transform showing the arrangement of spots corresponding to a SiC nanoparticle oriented to the $[011]$ zone axis along with the spots corresponding to the magnesium matrix oriented to the $[2\bar{1}\bar{1}0]$ zone axis. The $\{111\}$ planes of the SiC nanoparticle are parallel to the basal (0002) planes of the magnesium matrix. Panels **e** and **f** have been rotated 45° with respect to the original (**d**) for greater clarity.

9 μm diameter and 18 μm length. However, from the TEM analysis, we could not find any sign of twinning. Thus, the suppression of twinning in our samples is mainly due to the high density of nanoparticles rather than an effect of sample size.

In addition to the deformation by basal slip in grains oriented to the $[2\bar{1}\bar{1}0]$ zone axis, we also found activation of non-basal slip systems during microcompression of these larger polycrystalline micropillars (9 μm in diameter and 18 μm in length) cut from the as-solidified ingots, which probably contain a few grains. For example, pyramidal slip has been identified in a grain oriented to the $[01\bar{1}1]$ zone axis shown in the Extended Data Fig. 3. We believe that the strong hardening effect of the high density of nanoparticles on basal slip might enable the activation of non-basal slip systems. This is in agreement with a previous study showing that the ratio between the effective critical resolved shear stress for different slip systems may be close to unity when the additive hardening is high²⁴. Thus, a combined enhanced plasticity with high strength can be attained in our samples.

Therefore, the substantial strengthening and effective hardening of basal slip originating from the blocking of dislocations by the densely dispersed nanoparticles, along with effective load bearing enabled by well bonded Mg–SiC interfaces (Fig. 1d), sufficiently explains the remarkable strength observed in our material. An attempt to theoretically calculate the contribution of all strengthening mechanisms is discussed in Methods subsection ‘Strengthening mechanisms’.

To introduce the grain-boundary strengthening mechanism (the Hall–Petch effect), thermo-mechanical processing by high-pressure torsion (HPT) was applied to as-solidified samples to obtain a high density of grain boundaries. This allowed us to investigate the effect of uniformly dispersed nanoparticles on the grain refinement during HPT processing. Ten revolutions were applied to these samples to eliminate the influence of the initial grain size on the final grain size after HPT (process details are presented in Methods subsection

‘Fabrication of nanocomposites’). The results of microstructural characterization and microcompression tests of HPT samples are shown in Fig. 3. Dark-field TEM images were collected to measure the grain size of the samples without and with nanoparticles after HPT.

As an example, both dark-field TEM images in Fig. 3a and b were acquired by setting the objective aperture on the strongest magnesium rings of their corresponding selected area electron diffraction patterns. That means that grains highlighted in these dark-field TEM images are mainly oriented to the $\langle 10\bar{1}1 \rangle$ orientation, along with a few grains oriented to the $\langle 0002 \rangle$ and $\langle 10\bar{1}0 \rangle$ orientations. The average grain size obtained from at least 200 measured grains was about 105 ± 42 nm and 64 ± 40 nm for the Mg₂Zn and Mg₂Zn (14 vol% SiC) samples, respectively, as shown by the histograms in Fig. 3c and d. These results indicate slightly finer grain size in the samples with nanoparticles. The engineering stress–strain curves of HPT-processed samples and the morphology of post-deformed micropillars are shown in Fig. 3e–g. An additional strength enhancement of about 300 MPa is achieved by the HPT, slightly higher than the additional enhancement of 286 MPa for the samples without nanoparticles. The combined strengthening by dispersed nanoparticles and grain refinement after HPT can explain the remarkable yield strength of 710 ± 35 MPa attained for the Mg₂Zn (14 vol% SiC) sample, which is the highest yield strength reported for magnesium alloys and their composites (to the best of our knowledge)^{25,26}. Further discussion of the strengthening mechanisms involved can be found in Methods subsection ‘Strengthening mechanisms’.

Furthermore, the strong interfacial bonding between SiC nanoparticles and the magnesium matrix also leads to a substantial enhancement of Young’s modulus (tested by microindentation and discussed in Methods subsection ‘Enhancement of Young’s modulus’). Whereas the Young’s modulus for the Mg₂Zn sample is around 44 ± 5 GPa, the Young’s modulus for the Mg₂Zn (14 vol% SiC) sample increases up to

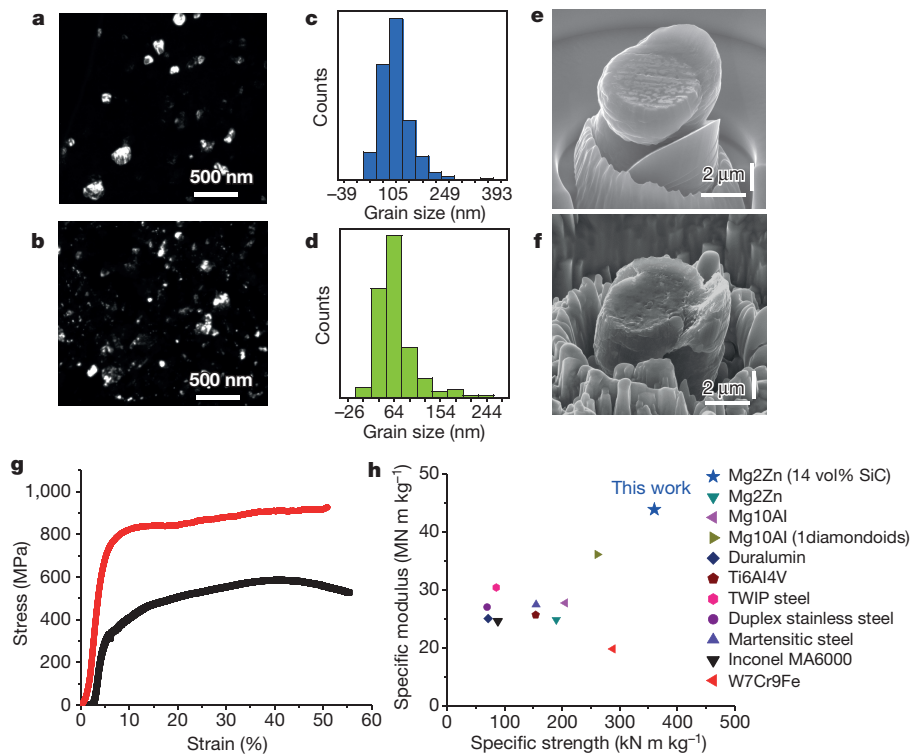


Figure 3 | Structure refinement and strength enhancement by HPT. **a, b,** Dark-field TEM images displaying the nanocrystalline grains of HPT-processed samples without (a) and with (b) nanoparticles. **c, d,** Histograms indicating the grain size distribution in both samples without (a) and with (b) nanoparticles. **e, f,** SEM images showing the morphology of post-deformed micropillars without (e) and with (f) nanoparticles. **g,** Engineering

stress-strain curves of HPT-processed magnesium samples without (black) and with (red) SiC nanoparticles. **h,** Specific modulus versus specific yield strength of HPT-processed Mg₂Zn (14 vol% SiC) in comparison with the results from micropillar testing of other metals and alloys. The nature and source of the data points are presented in Methods subsection 'Comparison with representative engineering alloys'.

86 ± 5 GPa. As a result, the Mg₂Zn (14 vol% SiC) sample after HPT not only exhibits the highest specific strength but also the highest specific modulus of all the reported data from micropillars of different metals and alloys (with diameters of 3.5–5 μm, to be consistent with our geometry) collected in Fig. 3h. The references for all the data points in Fig. 3h can be found in Methods subsection 'Comparison with representative engineering alloys'.

To prove that there is no size effect in the HPT-processed Mg₂Zn (14 vol% SiC) micropillars, microcompression tests on larger micropillars (9 μm in diameter and 18 μm in length) were also performed. A yield strength of 716 ± 38 MPa was obtained for these 9-μm micropillars, which is essentially the same as that for the 4-μm micropillars (710 ± 35 MPa). These results confirm that the strengths for 4-μm and 9-μm micropillars are almost equivalent, which is in good agreement

with the 'no size-induced strengthening' recently reported for magnesium micropillars bigger than 3.5 μm (ref. 19). Thus, we expect a similar high strength for the bulk samples. Further details can be found in Methods subsection 'Strengthening mechanisms'.

As is well known in the literature, improving the strength of magnesium at high temperatures is always a challenge^{27,28}. Precipitates obtained after heat treatment that would contribute to the strengthening tend to dissolve or grow at elevated temperatures, leading to a loss of strengthening¹⁴. To evaluate the high-temperature properties of the Mg₂Zn (14 vol% SiC) samples, micropillar compression tests were conducted at 200 °C, 300 °C and 400 °C inside a SEM chamber (details in Methods subsection 'Mechanical characterization').

As an example, Fig. 4a shows the stress-strain curve of an Mg₂Zn (14 vol% SiC) micropillar at 400 °C. A yield strength of about

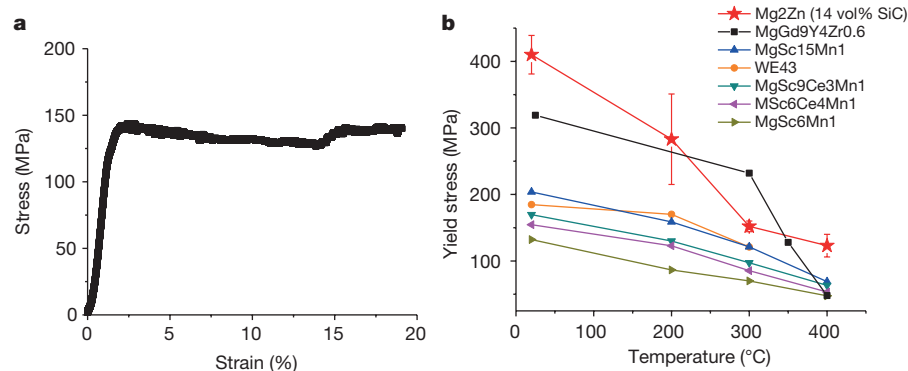


Figure 4 | Mechanical behaviour of as-solidified sample at elevated temperatures. **a,** Engineering stress-strain curve of Mg₂Zn (14 vol% SiC) micropillars at 400 °C. **b,** Yielding stress of Mg₂Zn (14 vol% SiC) at

elevated temperatures compared with other high-temperature magnesium alloys. Error bars represent s.d. of at least three data sets.

123 ± 17 MPa was obtained, which is about twice as much as that of the most heat-resistant magnesium alloys reported to date. Our results at different temperatures are compared with those from high-temperature magnesium alloys in Fig. 4b^{29,30}. Our Mg₂Zn (14 vol% SiC) samples offer a remarkable high-temperature strength.

In summary, our study on the dispersion and self-stabilization of SiC nanoparticles in molten magnesium suggests a new way of dispersing dense nanoparticles in metal matrices to achieve simultaneously enhanced strength, elastic modulus, plasticity and high-temperature stability. Ultrahigh-performance lightweight metals would improve energy efficiency and system performance in numerous applications. Although the method reported here is scalable in principle, many efforts are still needed to realize large-volume manufacturing for practical applications.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 April; accepted 4 November 2015.

- Pollock, T. M. Weight loss with magnesium alloys. *Science* **328**, 986–987 (2010).
- Lu, K. The future of metals. *Science* **328**, 319–320 (2010).
- Nie, J. F., Zhu, Y. M., Liu, J. Z. & Fang, X. Y. Periodic segregation of solute atoms in fully coherent twin boundaries. *Science* **340**, 957–960 (2013).
- Erbel, R. et al. Temporary scaffolding of coronary arteries with bioabsorbable magnesium stents: a prospective, non-randomised multicentre trial. *Lancet* **369**, 1869–1875 (2007).
- Knochel, P. A flash of magnesium. *Nature Chem.* **1**, 740 (2009).
- Nie, J.-F. Precipitation and hardening in magnesium alloys. *Metall. Mater. Trans. A* **43**, 3891–3939 (2012).
- Mortensen, A. & Llorca, J. Metal Matrix Composites. *Annu. Rev. Mater. Res.* **40**, 243–270 (2010).
- Liu, G. et al. Nanostructured high-strength molybdenum alloys with unprecedented tensile ductility. *Nature Mater.* **12**, 344–350 (2013).
- Zhang, Z. & Chen, D. Consideration of Orowan strengthening effect in particulate-reinforced metal matrix nanocomposites: a model for predicting their yield strength. *Scr. Mater.* **54**, 1321–1326 (2006).
- Ferguson, J. B., Sheykhi-Jaberi, F., Kim, C. S., Rohatgi, P. K. & Cho, K. On the strength and strain to failure in particle-reinforced magnesium metal-matrix nanocomposites (Mg MMNCs). *Mater. Sci. Eng. A* **558**, 193–204 (2012).
- Tjong, S. C. Novel nanoparticle-reinforced metal matrix composites with enhanced mechanical properties. *Adv. Eng. Mater.* **9**, 639–652 (2007).
- Chen, L. Y. et al. Novel nanoprocessing route for bulk graphene nanoplatelets reinforced metal matrix nanocomposites. *Scr. Mater.* **67**, 29–32 (2012).
- Ferkel, H. & Mordike, B. L. Magnesium strengthened by SiC nanoparticles. *Mater. Sci. Eng. A* **298**, 193–199 (2001).
- Sillekens, W. H. et al. The ExoMet project: EU/ESA research on high-performance light-metal alloys and nanocomposites. *Metall. Mater. Trans. A* **45**, 3349–3361 (2014).
- Xu, J. Q., Chen, L. Y., Choi, H. & Li, X. C. Theoretical study and pathways for nanoparticle capture during solidification of metal melt. *J. Phys. Condens. Matter* **24**, 255304 (2012).
- Chen, L.-Y., Peng, J.-Y., Xu, J.-Q., Choi, H. & Li, X.-C. Achieving uniform distribution and dispersion of a high percentage of nanoparticles in metal matrix nanocomposites by solidification processing. *Scr. Mater.* **69**, 634–637 (2013).
- Min, Y., Akbulut, M., Kristiansen, K., Golan, Y. & Israelachvili, J. The role of interparticle and external forces in nanoparticle assembly. *Nature Mater.* **7**, 527–538 (2008).
- Somekawa, H. & Schuh, C. A. Nanoindentation behavior and deformed microstructures in coarse-grained magnesium alloys. *Scr. Mater.* **68**, 416–419 (2013).
- Pozuelo, M., Chang, Y. W. & Yang, J. M. Effect of diamondoids on the microstructure and mechanical behavior of nanostructured Mg-matrix nanocomposites. *Mater. Sci. Eng. A* **633**, 200–208 (2015).
- Ye, J., Mishra, R. K., Sachdev, A. K. & Minor, A. M. In situ TEM compression testing of Mg and Mg-0.2 wt.% Ce single crystals. *Scr. Mater.* **64**, 292–295 (2011).
- Barnett, M. R., Keshavarz, Z., Beer, G. & Atwell, D. Influence of grain size on the compressive deformation of wrought Mg-3Al-1Zn. *Acta Mater.* **52**, 5093–5103 (2004).
- Stanford, N. & Barnett, M. R. Effect of particles on the formation of deformation twins in a magnesium-based alloy. *Mater. Sci. Eng. A* **516**, 226–234 (2009).
- Watanabe, H. Effect of second-phase precipitates on local elongation in extruded magnesium alloys. *J. Mater. Eng. Perform.* **22**, 3450–3454 (2013).
- Hutchinson, W. B. & Barnett, M. R. Effective values of critical resolved shear stress for slip in polycrystalline magnesium and other hcp metals. *Scr. Mater.* **63**, 737–740 (2010).
- Jian, W. W. et al. Ultrastrong Mg alloy via nano-spaced stacking faults. *Mater. Res. Lett.* **1**, 61–66 (2013).
- Inoue, A. et al. Novel hexagonal structure of ultra-high strength magnesium-based alloys. *Mater. Trans.* **43**, 580–584 (2002).
- Luo, A. A. Recent magnesium alloy development for elevated temperature applications. *Int. Mater. Rev.* **49**, 13–30 (2004).
- Luo, A. & Pekguleryuz, M. O. Cast magnesium alloys for elevated temperature applications. *J. Mater. Sci.* **29**, 5259–5271 (1994).
- Friedrich, H. E. & Mordike, B. L. *Magnesium Technology—Metallurgy, Design Data, Applications* (Springer, 2006).
- Xiao, Y., Zhang, X., Chen, J. & Jiang, H. Microstructures and mechanical properties of extruded Mg-9Gd-4Y-0.6Zr-T5 at elevated temperatures. *Chin. J. Nonfer. Met.* **16**, 709–714 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work is supported in part by the National Institute of Standards and Technology (NIST). We thank Y.-W. Chang, N. Bodzin and T. McLouth at the University of California, Los Angeles, for their help with FIB experiments, micropillar testing and elastic modulus measurements. We also thank C. Cao at the University of California, Los Angeles for his help with measuring the grain size of the as-solidified Mg₂Zn samples.

Author Contributions X.-C.L. and L.-Y.C. conceived the idea and designed the experiments. L.-Y.C. and H.C. fabricated the nanocomposites. X.-C.L. and J.-Q.X. developed the theoretical model for nanoparticle dispersion. X.M. conducted the high-pressure torsion experiment. L.-Y.C. and M.P. characterized the properties and microstructures. S.B. conducted micropillar compression testing at high temperature. L.-Y.C., X.-C.L., J.-Q.X., M.P. and S.M. analysed the data. L.-Y.C., X.-C.L., M.P. and S.M. wrote the paper. J.-M.Y. supervised M.P. for TEM characterization. S.M. supervised X.M. for the high-pressure torsion experiment. X.-C.L. supervised the whole work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to X.-C.L. (xccli@seas.ucla.edu).

METHODS

Fabrication of nanocomposites. Mg6Zn alloy was melted in an alumina crucible under the protection of CO₂ (99 vol%) and SF₆ (1 vol%). SiC nanoparticles were fed into the Mg6Zn alloy melt to 1.0 vol% and dispersed by ultrasonic processing with a frequency of 20 kHz and a peak-to-peak amplitude of 60 μm at 700 °C. After a slow solidification, a Mg6Zn (1 vol% SiC) ingot was obtained. To achieve a high-volume fraction of nanoparticles in the magnesium melt, SiC nanoparticles were concentrated by evaporating away magnesium and zinc from the Mg6Zn (1 vol% SiC) ingot (about 20 g) at 6 torr in a vacuum furnace. Then the samples (about 1.5 g) were cooled down slowly (with a cooling rate measured by a thermocouple of only 0.23 K s⁻¹) to room temperature inside the furnace.

To further enhance the strength of the materials, HPT was used as a secondary process. HPT, a popular method of severe plastic deformation, is effective in grain refinement of various materials and therefore elevates their strength based on the Hall–Petch relationship³¹. As-solidified magnesium alloys were punched into 10-mm-diameter disks. HPT was applied to each disk at room temperature with an imposed pressure of 1.0 GPa for ten revolutions at 1.5 r.p.m. to obtain sufficient grain refinement, to eliminate the influence of initial grain size on final grain size and to achieve more homogeneous deformation along the radial direction³¹.

Structural characterization. The distribution and dispersion of SiC nanoparticles were studied by means of SEM and TEM. To clearly reveal the nanoparticles, the SEM samples were first cleaned by low-angle ion milling (10°, to remove the nanometre-sized polishing powders) and then slightly etched by gallium ions (90°, to preferentially etch magnesium matrix) by FIB. The SEM images were acquired at a 52° tilt to expose the nanoparticles on the surface of the magnesium matrix. The composition of the materials was evaluated by energy-dispersive X-ray spectroscopy. The analysis was conducted with a FEI Nova 230 Variable Pressure SEM (VP-SEM) equipped with a Thermo Fisher Scientific energy-dispersive X-ray spectroscopy system at an accelerating voltage of 15 kV. The nanoparticles–magnesium matrix interfaces, the orientation of micropillars for mechanical testing, and the grain size in samples after high-pressure torsion were investigated by TEM. A FEI-Titan scanning TEM operated at 300 kV was used for this purpose. Thin-foil TEM samples were prepared by FIB.

Mechanical characterization. Microcompression tests were conducted at room temperature under displacement control mode and at a strain rate of 2×10^{-3} s⁻¹. A PI 85 SEM PicoIndenter (Hysitron Inc.) with a 5-μm flat punch diamond probe inside a FEI Nova 600 Nanolab Dual-Beam FIB-SEM was used for *in situ* experiments on 4-μm micropillars. An MTS Nanoindenter with a flat punch tip was used for microcompression testing on 9-μm-diameter micropillars. Micropillars of 4 μm and 9 μm in diameter (8 μm and 18 μm in length, respectively) were machined by FIB from the as-solidified samples with and without SiC nanoparticles and after HPT.

In situ quasi-static compression experiments at elevated temperatures were conducted using a PI 85 SEM PicoIndenter (Hysitron Inc.) with a 20-μm flat punch diamond probe inside a SEM (Versa 3D FIBSEM, FEI Company). TriboScan software (Hysitron Inc.) was used to monitor, capture and analyse the load–displacement data. The load–displacement data and the real-time video of deformation were synchronized and recorded during the experiment, which aided the post-experimental analysis.

In situ heating was conducted through the use of a resistive microelectromechanical systems (MEMS)-based heater which facilitates heating of a sample up to 450 °C (ref. 32). An integrated thin film of platinum on a quartz structure was used as heating element. Temperature was actively measured and feedback-controlled using an RTD (resistance temperature detector) sensor to ensure that the desired temperature is achieved and maintained within 0.1 °C. The sample was mounted using high-temperature conductive epoxy, EpoTek. Owing to the small size of the heating system, the region of elevated temperature is highly localized so that it minimizes extraneous heating of system components and provides the maximum level of stability for mechanical testing. To achieve thermal equilibrium between the probe and the sample, the probe was contacted with micropillars using a 10-μN load for 300 s before testing. In addition, thermal drift was also monitored and analysed for a preset time before each tests, and the measured drift rate was considered to correct the load–displacement data. Compression experiments at 200 °C, 300 °C and 400 °C were conducted on micropillars with a diameter of about 4 μm and length of about 8 μm using displacement control mode to a maximum strain of 25% with a strain rate of 2×10^{-3} s⁻¹.

Microindentation tests with an indent depth of 2 μm were performed to evaluate the elastic modulus from the unloading curves. An MTS Nanoindenter XP with a Berkovich tip was used. Vickers hardness measurements were made under loads of 4.9 N with a dwelling time of 10 s.

Nanoparticle self-stabilization mechanism. The self-stabilization of dispersed SiC nanoparticles in magnesium melt is attributed to a synergy of reduced van der Waals forces between the nanoparticles in molten magnesium, a high thermal

energy of the nanoparticles, and a high energy barrier preventing nanoparticle from sintering owing to a reasonable wettability between nanoparticles and molten magnesium, as schematically shown in Fig. 1c.

van der Waals attraction. For two SiC nanoparticles in magnesium melt at 1,000 K, the van der Waals interaction can be approximately estimated by the following equation³³:

$$W_{\text{vdw}}(D) = - \frac{(\sqrt{A_{\text{SiC}}} - \sqrt{A_{\text{Mg}}})^2}{6D} \left(\frac{R_1 R_2}{R_1 + R_2} \right) \quad (1)$$

where D is the distance between two nanoparticles in nanometres, A_{SiC} and A_{Mg} are the Hamaker constants for the van der Waals interaction and are 248 zJ and 206 zJ for SiC and molten magnesium, respectively¹⁵. R_1 and R_2 are the radii of two nanoparticles. Thus the van der Waals interaction between two similar SiC with radii of R in molten magnesium is:

$$W_{\text{vdw}}(D) = - \frac{(\sqrt{248} - \sqrt{206})^2}{6D} \frac{R}{2} \quad (2)$$

Equation (2) is only effective when two SiC nanoparticles interact in the magnesium molten with D approximately larger than two atomic layers (that is, 0.4 nm). Therefore, the maximum attraction, $W_{\text{vdw}}(\text{min})(D)$ between two SiC nanoparticles in magnesium melt, is estimated to be -12.17 zJ when $D = 0.4$ nm. The low van der Waals attraction potential is due to the small difference between the Hamaker constant of SiC and that of molten magnesium. If D is smaller than two atomic layers (that is, 0.4 nm) of magnesium, the van der Waals interaction contributes to the interfacial energy of the system together with much stronger interfacial chemical bonds^{15,33}.

Thermal energy for nanoparticle dispersion. The thermal energy of nanoparticles for Brownian motion, E_b , can be calculated by:

$$E_b = kT$$

where k is the Boltzmann constant and T is the absolute temperature. At the processing temperature of 1,000 K, E_b is 13.8 zJ, which is larger than the maximum van der Waals attraction in our Mg–SiC system. Therefore, driven by the thermal energy, SiC nanoparticles can break free from the van der Waals attraction in magnesium melt.

Energy barrier preventing nanoparticle contacting and sintering. At a high temperature, nanoparticles may sinter together if they are in contact, driven by a substantial drop of interfacial energy. In the Mg–SiC system, when two nanoparticles approach each other to a distance $D = 0.2$ nm, the last atomic layer of magnesium will be squeezed out. The Mg–SiC interface will then be replaced by SiC surfaces. The interfacial energy increase will be $W_{\text{barrier}} = S(\sigma_{\text{SiC}} - \sigma_{\text{SiC-Mg}}) = S\sigma_{\text{Mg}}\cos\theta$, where S is the effective area, σ_{SiC} is the surface energy of SiC, $\sigma_{\text{SiC-Mg}}$ is the interfacial energy between SiC and magnesium melt, σ_{Mg} is the surface tension of magnesium melt, and θ is the contact angle of magnesium melt on SiC surface. This equation clearly suggests that the better the wetting between nanoparticles and molten metal (smaller θ), the higher the energy barrier that prevents the nanoparticles contacting each other.

From the literature, the surface energy of liquid magnesium is 0.599 J m⁻² (ref. 34) and the surface energy of SiC is 1.45 J m⁻² (ref. 35). The contact angle is 83° (ref. 36). Hence, the interfacial energy between liquid magnesium and SiC will be 0.422 J m⁻² according to Young's equation. According to the Langbein approximation, the effective interaction area of two spheres is $S = \pi R D_0$ (where $D_0 = 0.2$ nm)³³. For two SiC nanoparticles with a diameter of 60 nm, the interfacial energy increase will be 3.87×10^4 zJ, which is more than 2,000 times higher than the thermal energy for Brownian motion. Thus, SiC nanoparticles will have little chance to overcome the energy barrier to contact each other for sintering. Therefore, the dispersed SiC nanoparticles in magnesium melt will be stabilized.

Strengthening mechanisms. Considering that no fine intermetallic precipitates were observed in our samples, the strengthening mechanism must be mainly caused by the densely dispersed nanoparticles. We believe that the absence of fine precipitates might be attributed to two factors. First, the solidification process we implemented directly cooled the alloy from the liquid state, and no posterior precipitation heat treatments were performed. Second, the low concentration of zinc is only 0.4 wt% higher than the equilibrium solubility limit of 1.6 wt%, and thus it is expected that zinc is mostly in solid solution in the as-solidified sample.

Previous studies show that the possible strengthening mechanisms in metal matrix nanocomposites include Orowan strengthening, increased dislocation density due to mismatch of thermal expansion coefficient, load bearing and Hall–Petch mechanisms. The potential strengthening mechanisms involved in the as-solidified and HPT-processed samples are discussed below.

Strengthening mechanism in the as-solidified sample. The total strengthening induced by dense dispersed nanoparticles in the as-solidified samples is about 360 MPa. Since the as-solidified samples we tested are single-crystal and with the same sample size for those with and without nanoparticles, there is no need to consider Hall–Petch strengthening. Additionally, from the TEM analysis we did not observe a large increase of dislocation density around the nanoparticles. This may be because the nanoparticles are too small to generate enough strain to induce a high density of dislocation around nanoparticles or the dislocations may be annealed out during the very slow cooling inside the furnace. Thus, the contribution to the strengthening from the increased dislocation density due to the mismatch of thermal expansion coefficient can be neglected.

The contribution by the Orowan strengthening ($\Delta\sigma_{\text{Orowan}}$) mechanism induced by well dispersed particles can be calculated by the following equation⁶:

$$\Delta\sigma_{\text{Orowan}} = \frac{\varphi G_m b}{d_p} \left(\frac{6V_p}{\pi} \right)^{1/3} \quad (3)$$

where G_m , b , V_p and d_p are the shear modulus of the matrix, the Burgers vector, the volume fraction and the size of the nanoparticles, respectively. φ is a constant equal to 2. Considering that in this study $G_m = 16.4$ GPa, $b = 0.32$ nm, $V_p = 0.14$ and $d_p = 60$ nm, the calculated $\Delta\sigma_{\text{Orowan}}$ is 113 MPa.

It is highly likely that the rest of the strengthening contribution is due to the load-bearing mechanism. The increase in strength due to load bearing can be calculated by the following equation³⁷:

$$\Delta\sigma_{\text{load}} = 1.5V_p\sigma_i$$

where σ_i is the interfacial bonding strength. To obtain a strengthening contribution of 262 MPa, the interfacial strength σ_i should be around 1,250 MPa. This result suggests a very strong interfacial bonding between SiC nanoparticles and the magnesium matrix, as we observed from the TEM analysis in Fig. 1d.

Strengthening mechanism in the HPT-processed sample. After HPT, we achieved an additional increment in strength for samples with and without nanoparticles of 300 MPa and 280 MPa, respectively. Since this secondary processing leads to polycrystalline materials with an average grain size of 105 ± 42 nm and 64 ± 40 nm for the Mg2Zn and Mg2Zn (14 vol% SiC) samples, respectively, the main contribution to the strength from HPT processing will be due to the Hall–Petch mechanism. The increased yield strength from the Hall–Petch effect can be calculated by the following equation:

$$\Delta\sigma_y = kd^{-1/2}$$

where d is the grain size and k is a constant. As expected from the Hall–Petch equation, the sample with the smallest grain size (Mg2Zn (14 vol% SiC)) results in the highest increment in strength.

A recent study shows that Hall–Petch strengthening may break down at a grain size of around 100 nm owing to grain-boundary rotation³⁸. However, in this work we report a further strengthening when the grain size is smaller than 100 nm. We believe that this might be due to the effect of ceramic nanoparticles on preventing grain-boundary rotation or that the critical grain size for when Hall–Petch strengthening breaks down in this alloy is below 60 nm. This is an interesting issue for further study.

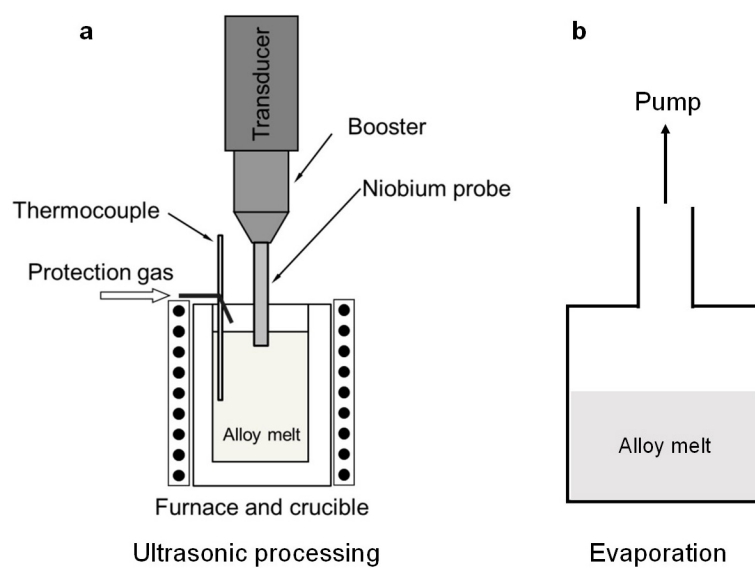
No size-induced strengthening. In our study, micropillar diameter was selected to be about 4 μm in order to avoid any size effect on the strengthening. It has recently been shown (ref. 19) that a strong size effect is observed in the strength of nanostructured Mg–Al micropillars with diameters smaller than 3.5 μm ; in contrast, no size effect was observed when diameters were bigger than 3.5 μm (ref. 19). In fact, the compressive strength of micropillars with diameters bigger than 3.5 μm was

similar to that of the bulk Mg–Al alloy and composite samples¹⁹. To further confirm that there is no size effect in our samples, we conducted compression tests on larger micropillars about 9 μm in diameter. These results show that the yield strength for 9- μm micropillars is 716 ± 38 MPa, which is nearly identical to the yield strength of 710 ± 35 MPa for 4- μm micropillars. This confirms that the strengths of 4- μm and 9- μm micropillars are essentially equivalent.

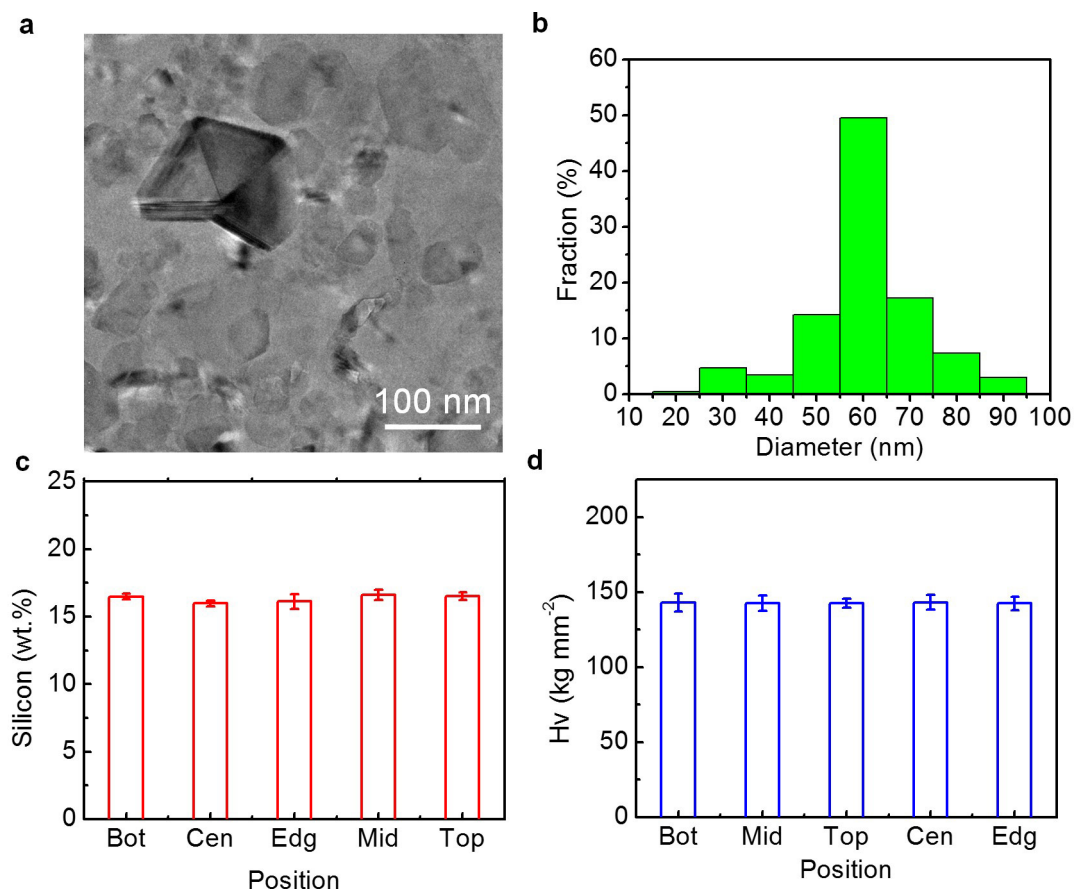
Enhancement of Young's modulus. The densely dispersed SiC nanoparticles also enable a substantial enhancement of Young's modulus from 44 ± 5 GPa in Mg2Zn to 86 ± 5 GPa in Mg2Zn (14 vol% SiC) samples. We believe that the increase in Young's modulus is due to the high Young's modulus of SiC (450 GPa) and the effective load bearing by the nanoparticles. The Young's modulus calculated by the rule of mixture is about 100 GPa in the Mg2Zn (14 vol% SiC) sample, which is in good agreement with the value tested by microindentation test (86 ± 5 GPa).

Comparison with representative engineering alloys. To compare our results fairly with the mechanical properties of representative engineering alloys, we have collected micropillar testing data for representative engineering metals with diameters of 3.5–5 μm , as shown in Fig. 3h. All data points with their corresponding references are listed as follows: Mg10Al (ref. 19), Mg10Al (1 diamondoids)¹⁹, Al4Cu1.3MgAg0.6Mn (Duralumin)³⁹, Ti6Al4V (ref. 40), Fe22Mn0.6C TWIP (twinning-induced plasticity) steel⁴¹, duplex stainless steel⁴², low-carbon martensitic steel⁴³, nickel superalloy (Inconel MA6000)⁴⁴, and W7Cr9Fe (ref. 45). The composition for W7Cr9Fe is in atomic per cent; for other alloys in weight per cent.

31. Zhilyaev, A. P. & Langdon, T. G. Using high-pressure torsion for metal processing: fundamentals and applications. *Prog. Mater. Sci.* **53**, 893–979 (2008).
32. Oh, Y., Asif, S. A. S., Cyrankowski, E. & Warren, O. L. Microelectro-mechanical heater. US patent WO2011066018 A1 (2011).
33. Israelachvili, J. N. *Intermolecular and Surface Forces* 211 (Academic Press, 2011).
34. Hashim, J., Looney, L. & Hashmi, M. S. J. The wettability of SiC particles by molten aluminium alloy. *J. Mater. Process. Technol.* **119**, 324–328 (2001).
35. Eustathopoulos, N., Nicholas, M. G. & Drevet, B. *Wettability at High Temperatures* 396 (Pergamon, 1999).
36. Zhang, D., Shen, P., Shi, L. X., Lin, Q. L. & Jiang, Q. C. Wetting and evaporation behaviors of molten Mg on partially oxidized SiC substrates. *Appl. Surf. Sci.* **256**, 7043–7047 (2010).
37. Nardone, V. C. & Prew, K. M. On the strength of discontinuous silicon carbide reinforced aluminum composites. *Scr. Metall.* **20**, 43–48 (1986).
38. Choi, H. J., Kim, Y., Shin, J. H. & Bae, D. H. Deformation behavior of magnesium in the grain size spectrum from nano- to micrometer. *Mater. Sci. Eng. A* **527**, 1565–1570 (2010).
39. Gu, R. & Ngan, H. W. Size effect on the deformation behavior of duralumin micropillars. *Scr. Mater.* **68**, 861–864 (2013).
40. Gong, J. & Wilkinson, A. J. A microcantilever investigation of size effect, solid-solution strengthening and second-phase strengthening for (a) prism slip in alpha-Ti. *Acta Mater.* **59**, 5970–5981 (2011).
41. Choi, W. S., De Cooman, B. C., Sandlöbes, S. & Raabe, D. Size and orientation effects in partial dislocation-mediated deformation of twinning-induced plasticity steel micro-pillars. *Acta Mater.* **98**, 391–404 (2015).
42. Guo, E. Y. *et al.* Mechanical characterization of microconstituents in a cast duplex stainless steel by micropillar compression. *Mater. Sci. Eng. A* **598**, 98–105 (2014).
43. Chen, P. *et al.* Microscale-calibrated modeling of the deformation response of dual-phase steels. *Acta Mater.* **65**, 133–149 (2014).
44. Girault, B., Schneider, A. S., Prick, C. P. & Arzt, E. Strength effects in micropillars of a dispersion strengthened superalloy. *Adv. Eng. Mater.* **12**, 385–388 (2010).
45. Cordero, Z. C. *et al.* Powder-route synthesis and mechanical testing of ultrafine grain tungsten alloys. *Metall. Mater. Trans. A* **45**, 3609–3618 (2014).

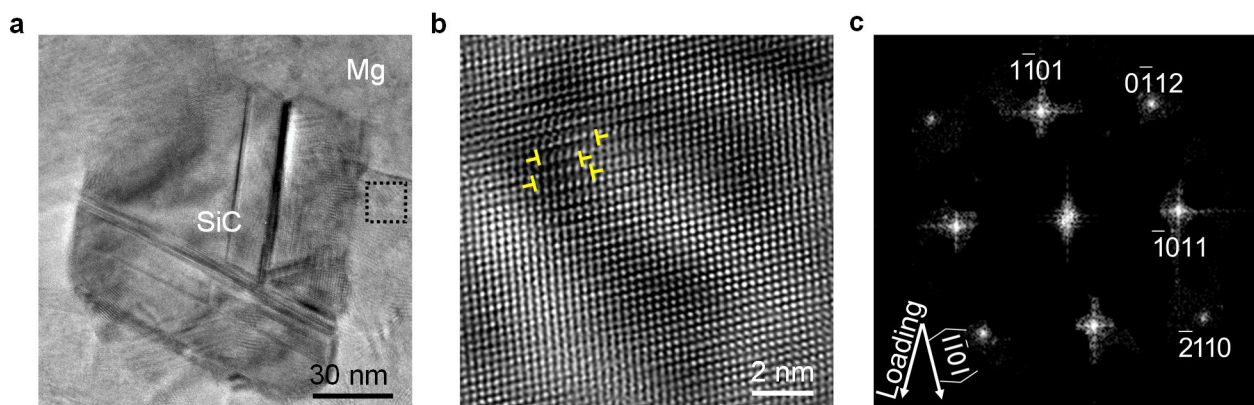


Extended Data Figure 1 | Fabrication of nanocomposites. **a**, Ultrasonic processing for nanoparticle feeding and dispersion. **b**, Vacuum evaporation for concentrating nanoparticles in magnesium.



Extended Data Figure 2 | Uniform distribution of nanoparticles across the whole sample. **a**, Bright-field TEM image showing the dispersed SiC nanoparticles in the magnesium matrix. **b**, A histogram indicating the SiC nanoparticle size distribution. **c**, **d**, Plots representing the amount of

Si (wt%; **c**) and Vickers microhardness (Hv; **d**) as a function of the position in the sample (bottom, middle, top, centre and edge). Error bars represent s.d. of six data sets in **d** and three data sets in **c**.



Extended Data Figure 3 | TEM analysis showing non-basal deformation mechanisms in a polycrystalline sample under microcompression.

a, Bright-field TEM image showing a SiC nanoparticle embedded in the magnesium matrix. **b**, High-resolution TEM image from the region

highlighted in **a** showing dislocations (indicated in yellow) terminated at stacking faults on the pyramidal $\{10\bar{1}1\}$ planes in a grain oriented to the $[01\bar{1}1]$ zone axis as indicated by its fast Fourier transform in **c**. The angle between the loading and pyramidal $\langle 1\bar{1}01 \rangle$ directions is around 30° .

Thermal vesiculation during volcanic eruptions

Yan Lavallée¹, Donald B. Dingwell², Jeffrey B. Johnson³, Corrado Cimarelli², Adrian J. Hornby¹, Jackie E. Kendrick¹, Felix W. von Aulock¹, Ben M. Kennedy⁴, Benjamin J. Andrews⁵, Fabian B. Wadsworth^{1,2}, Emma Rhodes⁴ & Gustavo Chigna⁶

Terrestrial volcanic eruptions are the consequence of magmas ascending to the surface of the Earth. This ascent is driven by buoyancy forces, which are enhanced by bubble nucleation and growth (vesiculation) that reduce the density of magma¹. The development of vesicularity also greatly reduces the ‘strength’ of magma², a material parameter controlling fragmentation and thus the explosive potential of the liquid rock³. The development of vesicularity in magmas has until now been viewed (both thermodynamically and kinetically) in terms of the pressure dependence of the solubility of water in the magma, and its role in driving gas saturation, exsolution and expansion during decompression. In contrast, the possible effects of the well documented negative temperature dependence of solubility of water in magma has largely been ignored. Recently, petrological constraints have demonstrated that considerable heating of magma may indeed be a common result of the latent heat of crystallization⁴ as well as viscous^{5,6} and frictional⁷ heating in areas of strain localization. Here we present field and experimental observations of magma vesiculation and fragmentation resulting from heating (rather than decompression). Textural analysis of volcanic ash from Santiaguito volcano in Guatemala reveals the presence of chemically heterogeneous filaments hosting micrometre-scale vesicles. The textures mirror those developed by disequilibrium melting induced via rapid heating during fault friction experiments, demonstrating that friction can generate sufficient heat to induce melting and vesiculation of hydrated silicic magma. Consideration of the experimentally determined temperature and pressure dependence of water solubility in magma reveals that, for many ascent paths, exsolution may be more efficiently achieved by heating than by decompression. We conclude that the thermal path experienced by magma during ascent strongly controls degassing, vesiculation, magma strength and the effusive–explosive transition in volcanic eruptions.

Volcanic eruptions result from magma buoyancy, largely powered by volatile exsolution. In standard models of magma ascent this exsolution is triggered by decompression^{8,9}. Upon ascent, gas bubbles (vesicles) expand and pressure build-up may precipitate fragmentation and explosive eruption¹. Yet the solubility, which sets the thermodynamic driving force for saturation and vesiculation in a volatile component has long been known to be a function of temperature as well¹⁰. Thus temperature changes may also generate magma vesiculation. Despite this, to our knowledge, no models of volcanic eruptions have explored the role of temperature in generating magmatic vesicularity.

The thermal evolution of magma in volcanic conduits has received increased attention in recent years. First, petrological studies have demonstrated that crystallizing magmas can heat up considerably (up to about 100 °C) owing to the latent heat liberated⁴—a process acting across the entire magmatic column. Second, zones in which magma undergoes strain localization during ascent also exhibit evidence of

considerable heating (up to about 250 °C) resulting from viscous energy dissipation^{5,6,11,12}. Third, the discovery of pseudotachylytes (caused by frictional melting during faulting) in erupted dome rocks¹³ and at the margin of lava spines⁷ indicates that fault friction can be an important contributor to the thermal budget of magma (locally up to about 1,000 °C), thus strongly affecting volcanic eruption dynamics¹⁴.

Evidence is mounting that magma ascent may often be controlled by strain localization near conduit margins¹⁵. Such strain localization in magmas has been proposed as a scenario leading to failure and potentially serving as a trigger for explosive eruptions^{16,17}. Careful examination of shallow volcanic conduit structures lends support to these proposals¹⁸. Magmatic conduits or dykes are relatively narrow (tens of centimetres to a few metres) at depths of a few kilometres¹⁹, so regions of strain localization may represent an important mass fraction of ascending magma. At shallow depths, where conduits can be wider (metres to a few tens of metres), areas of strain localization may not appear to be inevitable, yet the observation that shallow magma bodies are heavily fractured²⁰, and the influence of such fractures on surficial magma behaviour²¹ suggest that strain localization and its associated heat may play a large part throughout the length of the magmatic column.

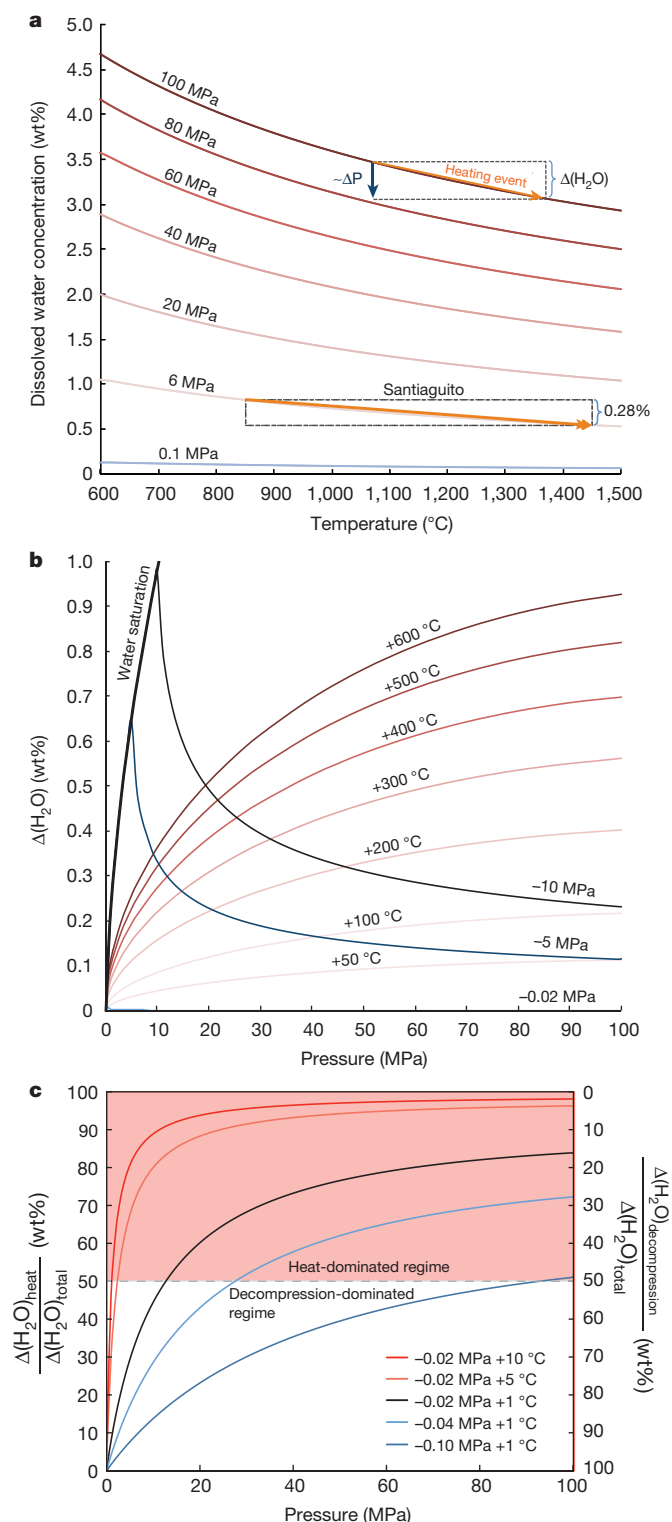
Estimates of magma ascent rates vary widely. In general, explosive eruptions have been associated with high ascent rates, reaching up to a few metres per second before fragmentation⁹. During such rapid ascent, magma decompresses (one metre per second corresponds to a decompression of 0.02 MPa per second) and simultaneously heat is generated in all areas where strain is localized, either by fault friction¹⁴ or viscous dissipation¹². The material record of such heat may or may not be documented in the products of the subsequent volcanic explosions. The mineralogical assemblage can often preserve information related to such heating^{7,13}, but if sufficient time passes then the assemblage will recover in response to the mean temperature and pressure conditions and evidence of fluctuations may be lost. The glassy state itself does not provide direct information from above the glass transition temperature, yet, indirectly, evidence of energy dissipation has been inferred from the morphology of the porous network preserved in glassy volcanic products^{5,6}. The difficulty of preservation of evidence of heating in ascending magma, or of the temperature history, is probably a major reason for its neglect in eruption models.

Temperature and pressure both affect the solubility of water^{22–24} (the dominant volatile component of volcanic activity) in magma. For a calc-alkaline rhyolitic melt, the temperature- and pressure-dependence of water solubility can be estimated by²⁴:

$$(\text{H}_2\text{O})_{\text{total}} = \frac{354.94P^{0.5} + 9.623P - 1.5223P^{1.5}}{T} + 0.0012439P^{1.5} \quad (1)$$

where $(\text{H}_2\text{O})_{\text{total}}$ is the total dissolved H_2O content (in weight per cent, wt%), T is temperature (in K), and P is pressure (in MPa). Figure 1a shows that owing to the strongly retrograde nature of the H_2O solubility curve at low pressures, an increase in temperature is a driving force

¹Department of Earth, Ocean and Ecological Sciences, University of Liverpool, Liverpool L69 3GP, UK. ²Department of Earth and Environmental Sciences, Ludwig Maximilian University of Munich, Theresienstrasse 41/III, 80333 Munich, Germany. ³Department of Geosciences, Boise State University, Boise, Idaho, USA. ⁴Geological Sciences, University of Canterbury, Private Bag 4800, 8140 Christchurch, New Zealand. ⁵Department of Mineral Sciences, Smithsonian Institution, Washington, District of Columbia, USA. ⁶Instituto Nacional de Sismología, Vulcanología, Meteorología, e Hidrología (INSIVUMEH), 7a Avenue 14-57, Zone 13, Guatemala City, Guatemala.



for vesiculation in this pressure range. This temperature dependence is clearly large enough to have a substantial effect on water saturation during magma ascent in conduits.

We analysed the potential magnitude of water exsolution $\Delta\text{H}_2\text{O}$ that is due to (1) decompression and (2) heating at a magmatic temperature of 850 °C (Fig. 1b). The comparison of the individual effects of decompression versus heating yields striking results. We found that events that heat magma by hundreds of degrees, as described above, strongly drive substantial exsolution and vesiculation. For an ascent rate of one metre per second (that is, 0.02 MPa per second, which is capable of triggering explosive events), 1 K of heating has the potential to generate more

Figure 1 | Water concentration in rhyolitic magmas.

a, Thermobarometric limits on water concentration²⁴ show that the heat induced by mechanical work (orange arrows) during magma ascent causes a decrease in water solubility, $\Delta(\text{H}_2\text{O})$. This decrease in concentration may be related to an equivalent decompression, $\sim\Delta P$. At Santiaguito, thermal input of, for example, ~ 600 °C owing to short-lived faulting events may reduce water solubility by 0.28 wt%. **b**, Water exsolution, $\Delta(\text{H}_2\text{O})$, driven by thermal input (red curves) versus decompression events (blue curves) for an ascending magma at a nominal temperature of 850 °C. These heating and decompression events are computed as a function of melt pressure at which the event initiates in the magmatic column. **c**, Fraction of the total water concentration exsolved from the action of heat (left y axis), versus that of decompression (right y axis) for different decompression and heating events. The data shows that thermal input (which acts on a timescale of seconds) generally induces more water exsolution than decompression.

water exsolution than 0.02 MPa of decompression from initial pressures greater than 13 MPa (Fig. 1c), and further heating can be the main driving force for vesiculation. Expressing it in a different way, a decompression event exceeding 0.1 MPa (>5 m of ascent) would be required to exsolve more water than that exsolved by 1 °C of heating. We therefore conclude from this analysis that the thermal path of decompressing magma can greatly influence volatile exsolution. It is thus easy to envisage scenarios of heating-dominated or ‘thermal’ vesiculation during magma ascent at moderate pressures, and below we provide evidence to support the assertion that such thermal exsolution is also dominant during strain localization in magma at shallower depths.

We have examined eruptive products at the Santiaguito dome complex. The active Caliente lava dome offers one of the most spectacular displays of cyclic, piston-like eruptive activity ever recorded, often climaxing in gas-and-ash explosions along concentric fractures^{21,25,26} (Fig. 2a). Proximal monitoring of this dome has revealed a regular (~ 26 min) periodicity in ground inflation–deflation cycles²⁷. At the expansion maxima, the propagation of arcuate faults across the dome’s surface is observed and the dome’s centre thrusts upward and collapses back, followed by dome deflation²¹. Gas-and-ash explosions occur episodically along the faults, coincident with very-long-period seismic events, which have been interpreted to be associated with gas flow in fractures at the inflation maximum (Fig. 2b)²⁷. In the analysis that follows, the rates of inflation and deflation during ash release and the magnitude and rate of slip are of central importance. Ash ejection occurs only during the fastest inflation–deflation cycles (Fig. 2b)²⁷. In these cases, the arcuate faults undergo a metre of uplift and collapse within one second, corresponding to a slip rate of $<2\text{ m s}^{-1}$ (ref. 21). Importantly, these lava dome dynamics leave striation and slickensides (frictional marks) on the blocks forming the dome carapace.

Textural examination of volcanic ash collected upon deposition in November 2012 and November 2014 provides several examples of the material consequences of such frictional processes (Extended Data Figs 1–4). The interstitial glass phase reveals a juxtaposition of chemically distinct mingled filaments with different shades of grey on back-scattered electron (BSE) images obtained by scanning electron microscopy (SEM; Fig. 2c; Extended Data Figs 3 and 4). The contacts between the light- and darker-toned filaments are diffuse and fluid (unlike crystals with sharp and angular boundaries). The very fine nature of these filaments and the diffuse boundaries prevent us from accurately using standard geochemical analysis techniques, but the greyscale values observed (which reflect the atomic number and thus chemical variations within and between phases) provide clear evidence of chemical heterogeneity (Fig. 2c; Extended Data Fig. 4). These melt phases have evidently mingled with the original interstitial melt on timescales insufficient for homogenization, presumably immediately before the fragmentation and eruption that locked in these dynamic features.

The mingling textures exhibited by the Caliente ash mirror those of protomelts resulting from selective melting of individual

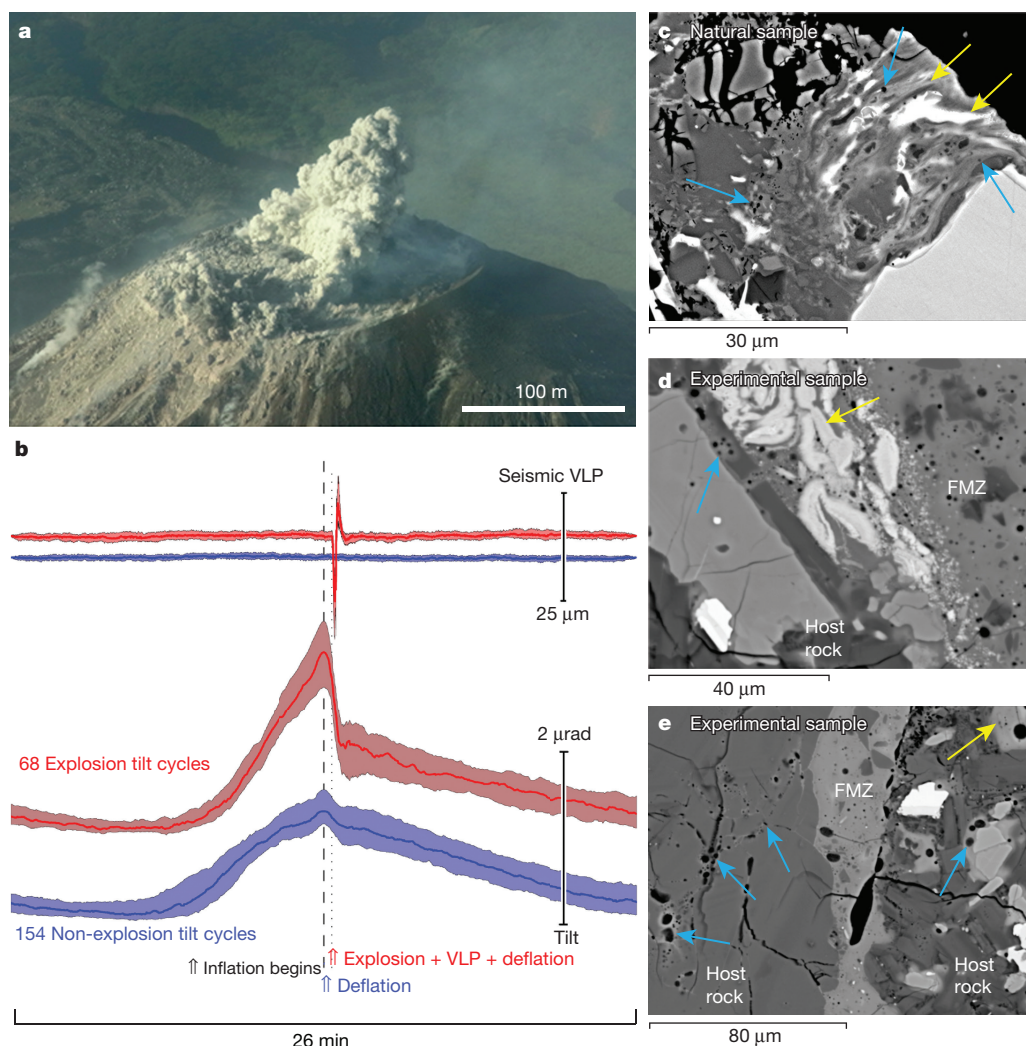


Figure 2 | Explosive eruptions caused by superheated vesiculation. **a**, Gas-and-ash explosion occurring along an arcuate fault on 10 November 2012 at Caliente dome, Santiaguito. **b**, Seismic signals (upper) and tilt data (lower) for explosive (red) and non-explosive (blue) inflation–deflation cycles associated with piston-like dynamics at Santiaguito²⁷. The solid lines display the average of the 26-min cycles over the five-day-long data set, whereas the shaded areas exhibit the spread in the data. Gas-and-ash explosion cycles differ markedly from non-explosive cycles and are characterized by faster and stronger inflation or deflation as well as very-long-period (VLP) seismic events. Note that the exact timing of the seismic and tilt signals may be offset slightly. **c**, BSE image showing heterogeneous protomelt filaments (yellow arrows) with different greyscale values (a proxy for chemical composition; darker grey indicates lighter elements and vice versa) extruded from crystals present in a volcanic ash particle sampled on 12 November 2012; some protomelts host vesicles (blue arrows). **d**, BSE image showing the shearing of protomelts near the main frictional melt zone (FMZ), produced experimentally by fault slip. **e**, BSE image showing vesiculation of the interstitial melt near the experimental fault zone, caused by high local temperature.

crystals that have been observed in the products of frictional melting experiments^{28,29}. Such experiments involve an extremely rapid heating rate (more than tens to hundreds of degrees Celsius per second) and therefore highly disequilibrium melting induced by fault friction^{28–30}. We propose here that the Caliente ash samples contain volcanic pseudotachylite; evidence of the syn-eruptive operation of frictional heating sufficient to generate melting in the piston-like events at Caliente dome. Notably, the protomelts present in the ash contain vesicles (as indicated by blue arrows on Fig. 2c). The crystalline phases present are anhydrous and thus cannot serve as a source of water for vesiculation, so we suggest that vesiculation took place in the interstitial melt. If so, these frictional melts contain clear evidence of thermal vesiculation in volcanic products.

As an experimental demonstration of the feasibility of thermal vesiculation, we have performed fault friction experiments under conditions designed to simulate the piston-like gas-and-ash explosion events at Caliente²¹. During the experiments the flat ends of two hollow, cylindrical cores of a Caliente dome rock were pushed together at an applied normal stress of 6 MPa (representative of the depth of tilt and seismic sources²⁷) and one core was rotated (against the other) at an equivalent velocity of 1 m s^{−1} (see Methods and Extended Data Fig. 5). Friction experiments on magmas have shown that under such conditions frictional melting takes place within as little as about 10 cm of slip^{13,14,28} confirming the feasibility of this process.

As noted above, microscopic inspection of the fault products experimentally generated in the Caliente dome rock reveals the presence of multiple, chemically heterogeneous melt filaments extruded from crystals adjacent to the fault zones (Fig. 2d; Extended Data Fig. 6).

In addition, the interstitial glass of the host rock in the first 0.3–0.4 mm near the fault zone has partially vesiculated (Fig. 2e; Extended Data Fig. 7). To ensure that vesiculation resulted from substantial heat near the fault zone, we have tested the stability of dissolved water in this dome rock at background magmatic temperature by subjecting two small cores to 850 °C for 30 min and 15 h, respectively. We observe that no water exsolved to form vesicles, even after a 15-h dwell (Extended Data Fig. 8). We conclude from these experiments that both the generation of crystal protomelts and the surrounding vesiculation result directly from the frictional work converted to substantial heat during faulting events, and are not due to residence at magmatic temperature. From the similarity of these experimental products of frictional melting to the natural samples of Caliente (described above) we deduce that the cyclic phenomena observed during dome extrusion and explosions at Caliente occur in the presence of strain localization, accompanied by thermal vesiculation.

The occurrence of superheated vesiculation at Caliente can be assessed by modelling the conversion of mechanical work to heat (ΔT) during friction, using³¹:

$$\Delta T = \frac{\mu \sigma_n V \sqrt{t}}{\rho C_p \sqrt{\pi k}} \quad (2)$$

Using Byerlee's friction coefficient μ of 0.85 (at static conditions), a normal stress σ_n of 6 MPa (ref. 27), a slip velocity V of 1 m s^{−1} for a duration t of 0.5 s (ref. 21), a density ρ of 2,630 kg m^{−3} (determined by helium pycnometry), a specific heat capacity C_p of 900 J kg^{−1} K^{−1}, and a thermal diffusivity k of 10^{−6} m² s^{−1}, uplift of the dome would generate a local temperature increase of 860 °C along the arcuate faults.

Given that the magma already resides at $\sim 850^\circ\text{C}$ (ref. 32), and that experimental work has shown that only moderate temperature increase occurs once frictional melt lubricates a slip zone^{13,14,28}, the temperature would not be expected to greatly exceed the melting temperatures of the main rock-forming minerals in the Caliente lava (labradorite and enstatite, which melt at $>1,300^\circ\text{C}$ and $>1,400^\circ\text{C}$, respectively³⁰). This magnitude of heating would induce water exsolution from the melt in zones of strain localization. Owing to the current eruptive cycles and outgassing activity at Caliente, we consider the system to be open to an extent that allows for exsolution of any oversaturated volatile fraction; thus a total of 0.83 wt% would be expected to remain in the magma at the point of fragmentation at 6 MPa (Fig. 1a). Heating of ~ 550 – 860°C would induce a dramatic oversaturation in water of 0.26–0.35 wt%. Faulting, creation of new surface area, and forced convection during frictional melting would all serve to minimize effective diffusion path lengths and enhance the completion of water exsolution. With such overheating, and thus heightened H_2O diffusivity, the kinetic limitation to vesiculation (nucleation and growth) should also be easily overcome, promoting foaming. At a depth of about 300 m such vesiculation would, in turn, reduce the strength of magma and thereby trigger fragmentation³³. We therefore conclude that vesiculation can be induced by rapid heating in the conduit.

Water is central to magma ascent dynamics and its contribution to magmatic and volcanic processes results from a combination of both pressure and temperature. Decompression is inevitable and acts throughout magma ascent. Here we argue that heating via both crystallization and shearing processes are equally inevitable. More specifically, the magnitude of viscous and frictional heating may be prodigious, and thus exert a primary control on volatile exsolution. At the rates and magnitudes of heating discussed here, the solubility of water in a melt should be affected before heat loss by thermal conductivity to the cooler surroundings—whether in the core of the magmatic column (where further water may exsolve) or in the country rock—could serve to counteract local heating. Heating during magma ascent deserves adequate consideration in conduit transport and eruption models.

The idea that temperature may dominate the dynamics of water saturation and vesiculation during magma transport in volcanic conduits means that the thermal path experienced by magmas during ascent need to be better constrained. A thorough reassessment of strain localization across deep dykes and shallow conduits should lead to the quantification of shear heating during magma transport. In light of the demonstration that heating may supercede decompression as a driving force for degassing, we call for this concept to be included in the simulation and analysis of magma ascent and eruption.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 January; accepted 6 October 2015.

- Sahagian, D. Magma fragmentation in eruptions. *Nature* **402**, 589 (1999).
- Vasseur, J., Wadsworth, F. B., Lavallée, Y., Hess, K.-U. & Dingwell, D. B. Volcanic sintering: timescales of viscous densification and strength recovery. *Geophys. Res. Lett.* **40**, 5658–5664 (2013).
- Dingwell, D. B. Volcanic dilemma: flow or blow? *Science* **273**, 1054–1055 (1996).
- Blundy, J., Cashman, K. & Humphreys, M. Magma heating by decompression-driven crystallization beneath andesite volcanoes. *Nature* **443**, 76–80 (2006).
- Rosi, M., Landi, P., Polacci, M., Di Muro, A. & Zandomenighi, D. Role of conduit shear on ascent of the crystal-rich magma feeding the 800-year-BP Plinian eruption of Quilotoa Volcano (Ecuador). *Bull. Volcanol.* **66**, 307–321 (2004).
- Wright, H. M. N. & Weinberg, R. F. Strain localization in vesicular magma: implications for rheology and fragmentation. *Geology* **37**, 1023–1026 (2009).
- Kendrick, J. E. *et al.* Extreme frictional processes in the volcanic conduit of Mount St. Helens (USA) during the 2004–2008 eruption. *J. Struct. Geol.* **38**, 61–76 (2012).
- Martel, C. & Schmidt, B. C. Decompression experiments as an insight into ascent rates of silicic magmas. *Contrib. Mineral. Petrol.* **144**, 397–415 (2003).
- Proussevitch, A. A. & Sahagian, D. L. Dynamics and energetics of bubble growth in magmas: Analytical formulation and numerical modeling. *J. Geophys. Res. Solid Earth* **103**, 18223–18251 (1998).

- Ghiorso, M. S. & Sack, R. O. Chemical mass-transfer in magmatic processes. 4. A revised and internally consistent thermodynamic model for the interpolation and extrapolation of liquid-solid equilibria in magmatic systems at elevated temperatures and pressures. *Contrib. Mineral. Petrol.* **119**, 197–212 (1995).
- Hess, K. U., Cordonnier, B., Lavallée, Y. & Dingwell, D. B. Viscous heating in rhyolite: an in situ determination. *Earth Planet. Sci. Lett.* **275**, 121–126 (2008).
- Mastin, L. G. The controlling effect of viscous dissipation on magma flow in silicic conduits. *J. Volcanol. Geotherm. Res.* **143**, 17–28 (2005).
- Kendrick, J. E. *et al.* Seismogenic frictional melting in the magmatic column. *Solid Earth* **5**, 199–208 (2014).
- Kendrick, J. E. *et al.* Volcanic drumbeat seismicity caused by stick-slip motion and magmatic frictional melting. *Nature Geosci.* **7**, 438–442 (2014).
- Lavallée, Y., Hess, K.-U., Cordonnier, B. & Dingwell, D. B. Non-Newtonian rheological law for highly crystalline dome lavas. *Geology* **35**, 843–846 (2007).
- Gonnermann, H. M. & Manga, M. Explosive volcanism may not be an inevitable consequence of magma fragmentation. *Nature* **426**, 432–435 (2003).
- Papale, P. Strain-induced magma fragmentation in explosive eruptions. *Nature* **397**, 425–428 (1999).
- Tuffen, H., Dingwell, D. B. & Pinkerton, H. Repeated fracture and healing of silicic magma generate flow banding and earthquakes? *Geology* **31**, 1089–1092 (2003).
- Noguchi, S., Toramaru, A. & Nakada, S. Groundmass crystallization in dacite dykes taken in Unzen scientific drilling project (USDP-4). *J. Volcanol. Geotherm. Res.* **175**, 71–81 (2008).
- Stasiuk, M. V. *et al.* Degassing during magma ascent in the Mule Creek vent (USA). *Bull. Volcanol.* **58**, 117–130 (1996).
- Johnson, J. B., Lees, J. M., Gerst, A., Sahagian, D. & Varley, N. Long-period earthquakes and co-eruptive dome inflation seen with particle image velocimetry. *Nature* **456**, 377–381 (2008).
- Holtz, F., Behrens, H., Dingwell, D. B. & Johannes, W. H_2O solubility in haplogranitic melts—compositional, pressure and temperature-dependence. *Am. Mineral.* **80**, 94–108 (1995).
- Ryan, A. G., Russell, J. K., Nichols, A. R. L., Hess, K.-U. & Porritt, L. A. Experiments and models on H_2O retrograde solubility in volcanic systems. *Am. Mineral.* **100**, 774–786 (2015).
- Liu, Y., Zhang, Y. X. & Behrens, H. Solubility of H_2O in rhyolitic melts at low pressures and a new empirical model for mixed H_2O – CO_2 solubility in rhyolitic melts. *J. Volcanol. Geotherm. Res.* **143**, 219–235 (2005).
- Holland, A. S. P., Watson, I. M., Phillips, J. C., Caricchi, L. & Dalton, M. P. Degassing processes during lava dome growth: insights from Santiaguito lava dome, Guatemala. *J. Volcanol. Geotherm. Res.* **202**, 153–166 (2011).
- Scharff, L., Hort, M. & Gerst, A. The dynamics of the dome at Santiaguito volcano, Guatemala. *Geophys. J. Int.* **197**, 1–17 (2014).
- Johnson, J. B., Lyons, J. J., Andrews, B. J. & Lees, J. M. Explosive dome eruptions modulated by periodic gas-driven inflation. *Geophys. Res. Lett.* **41**, 6689–6697 (2014).
- Hornby, A. J. *et al.* Spine growth and seismogenic faulting at Mt. Unzen, Japan. *J. Geophys. Res. Solid Earth* **120**, 4034–4054 (2015).
- Lin, A. M. & Shimamoto, T. Selective melting processes as inferred from experimentally generated pseudotachylytes. *J. Asian Earth Sci.* **16**, 533–545 (1998).
- Spray, J. G. in *Annual Review of Earth and Planetary Sciences* Vol. 38 (ed. Jeanloz, R. F. K. H.) 221–254 (2010).
- Carlsaw, H. S. & Jaeger, J. C. *Conduction of Heat in Solids* 2nd edn (Oxford Univ. Press, 1959).
- Harris, A. J. L. & Flynn, L. P. The thermal stealth flows of Santiaguito dome, Guatemala: implications for the cooling and emplacement of dacitic block-lava flows. *Geol. Soc. Am. Bull.* **114**, 533–546 (2002).
- Zhang, Y. X. A criterion for the fragmentation of bubbly magma based on brittle failure theory. *Nature* **402**, 648–650 (1999).

Acknowledgements We thank A. Pineda, the staff at the National Institute for Seismology, Volcanology, Meteorology and Hydrology of Guatemala (INSIVUMEH) and the Policía Nacional Civil de Guatemala for support with the field campaign. This work was supported by a European Research Council Starting Grant to Y.L. on ‘Strain Localisation in Magmas’ (SLiM, grant number 306488) and an Advanced Grant to D.B.D. on ‘Explosive volcanism in the Earth system’ (EVOKES, grant number 247076). J.B.J. acknowledges the National Science Foundation EAR-grant number 1151662. This work was partially funded by the European Union’s seventh programme for research, technological development, and demonstration under grant agreement 282759 (VUELCO), and by an AXA grant ‘Risk from Volcanic Ash in the Earth System’. We are grateful to K. Genereau and L. Mastin for constructive reviews.

Author Contributions Y.L., J.E.K., F.W.v.A., A.J.H., F.B.W., B.J.A., B.M.K. and D.B.D. conceptualized the model. G.C. facilitated fieldwork and supported data analysis. Y.L., C.C., F.B.W., A.J.H., B.M.K. and E.R. performed fieldwork, collected samples and Y.L., F.W.v.A., A.J.H. and J.E.K. analysed the ash. A.J.H., J.E.K. and Y.L. performed the experiments. J.B.J. performed the geophysical analysis. All authors contributed to the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.L. (yan.lavallee@liverpool.ac.uk).

METHODS

Volcanic ash sampling and analysis. The ash samples were collected after each explosion from a location (14° 44' 35.11" N, 91° 33' 40.69" W) approximately 275 m east-northeast from the active Caliente vent. The ash was collected by spreading a clean, 1.4 m × 1.4 m synthetic sheet. We used a paintbrush to carefully brush deposited ash into sample bags. The sheets were thoroughly cleaned after each sample collection and laid out to collect the ash of subsequent events. Owing to the proximity of the sampling location, we are very confident of the source and timing of the ash employed in this study.

The grain size of the sampled volcanic ash was measured using a laser diffraction particle size analyser from Coulter. The density was determined on 25-mm-diameter and 50-mm-long rock cores using a 100-cm³ helium pycnometer from Micromeritics.

SEM analysis and energy-dispersive X-ray spectroscopy. Geochemical mapping across the natural samples and the experimental products was conducted in a Phillips XL 30 SEM using BSE and energy-dispersive X-ray spectroscopy (EDS) run on the Oxford Instruments INCA software. BSE images provide an excellent means of identifying frictional melting textures, because the grey value of each phase relates to the atomic number, or the density of major elements representing the geochemical composition³⁴. A dense phase consisting of heavy elements elastically reflects more electrons and thus shows up in light grey on a BSE image; conversely, an elementally light phase shows up in dark grey.

EDS was used to map the chemical concentration of major elements present in the different phases observed by BSE imaging. The EDS system allows mapping of the distribution of these elements across the main phases. We used an electron beam of 5.5 µm at 20 keV and 8 nA. For the purpose of this study, we monitored the distribution of Si, Mg, Fe, Ti, Na and Al. Comparison of BSE and EDS images verify that the filaments have different chemical compositions.

Electron probe micro-analysis. Geochemical analysis of the phases present in the natural samples and the experimental products was performed in a CAMECA SX 100 Electron Probe Micro Analyser at the Ludwig Maximilian University of Munich in Germany. Probing of the glass and mineral phases was done using a focused electron beam of 15 keV and 20 nA (Extended Data Fig. 6). Note that because we used a focused beam on glass, the measured concentrations of the alkalis, namely Na and K, are reduced by some 0.1–0.3 wt% from what is likely to be present; however, the filaments were too thin to be measured with a defocused

beam, which would yield higher inaccuracy. Despite this, the results reveal the chemical distinction between the different phases.

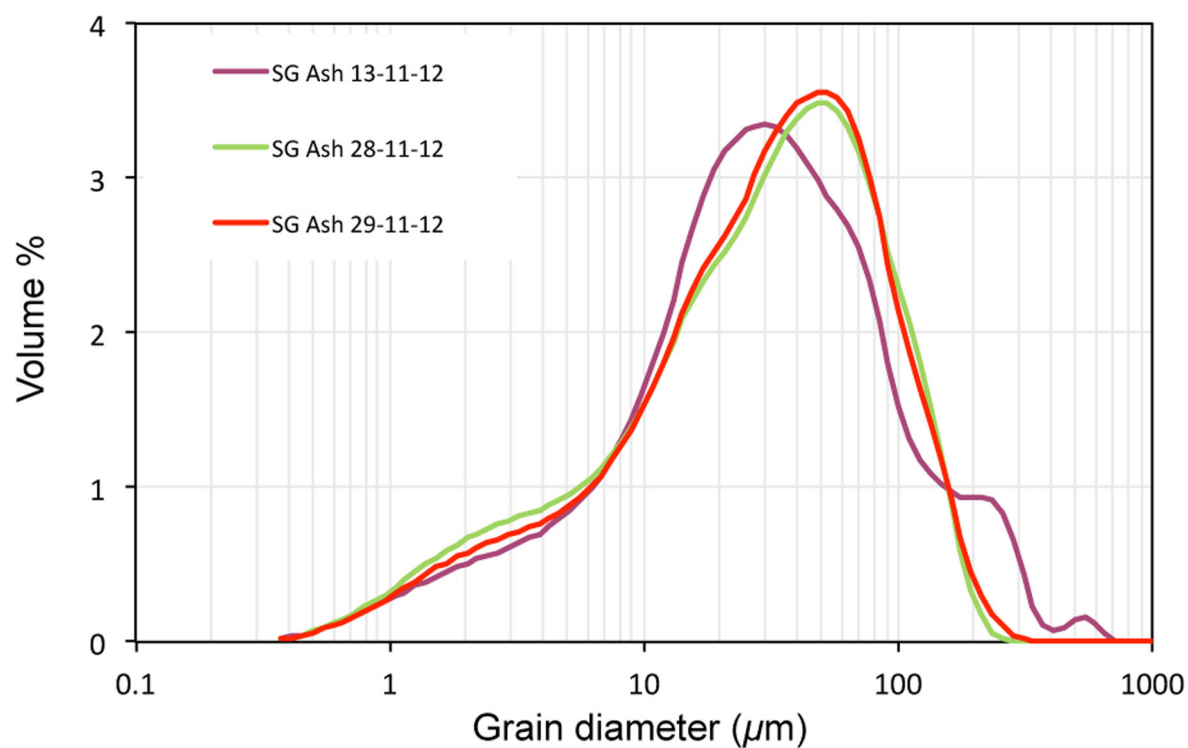
In Extended Data Fig. 6, we present the chemical composition of only the primary minerals and glass, and the protomelts and main frictional melt from the experiments, because the phases were large enough to be analysed. In the natural ash, the filaments are rarely larger than 1 µm (see Extended Data Figs 2–4) and so electron microprobe analysis was impracticable without a large degree of contamination from surrounding phases; hence, we used the greyscale in BSE images as well as EDS elemental maps to verify the occurrence of the same processes as observed in the experimental samples.

Fault slip experiments. The friction experiment was conducted in a low- to high-velocity rotary shear apparatus at the University of Liverpool, designed by T. Shimamoto and built by Marui, Japan. The experiment was conducted on two hollow, cylindrical samples with outer and inner diameters of 24.99 mm and 15.86 mm, respectively (Extended Data Fig. 5). The samples were axially loaded using an air actuator at a normal stress of 6.0 MPa, as constrained by the depth of seismicity, and slip was applied on one rotating sample via a servo motor operated at 1,200 rotations per minute, to induce an equivalent slip rate of 1 m s⁻¹, while the other sample was held stationary (see Hirose and Shimamoto³⁵ for further detail of apparatus and method). After the test, the sample was cut and a thin section was prepared.

Testing the stability of volatiles in the dome rock at eruptive temperature. We conducted complementary experiments to test the ability of the rock to vesiculate at high temperature to ensure that foaming observed in the friction experiments results from the very high temperatures achieved during fault slip, instead of simply because the rock used contains a concentration of water (quenched-in at high pressure) higher than that which is stable at atmospheric pressure. For this purpose, two small 8 mm × 8 mm cylindrical samples were heated to a magmatic temperature of 850 °C and one was allowed to dwell for 30 min while the other was allowed to dwell for 15 h. After the experiment, the samples were cut, polished and carbon-coated for SEM analysis.

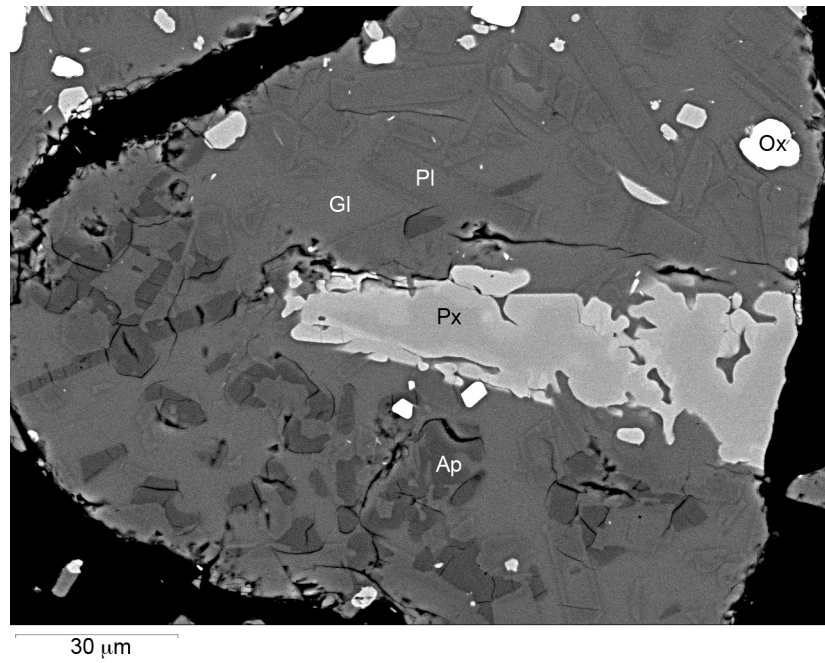
34. Petruk, W. *Applied Mineralogy in the Mining Industry* 1st edn, 268 (Elsevier, 1990).

35. Hirose, T. & Shimamoto, T. Growth of molten zone as a mechanism of slip weakening of simulated faults in gabbro during frictional melting. *J. Geophys. Res. Solid Earth* **110**, <http://dx.doi.org/10.1029/2004JB003207> (2005).



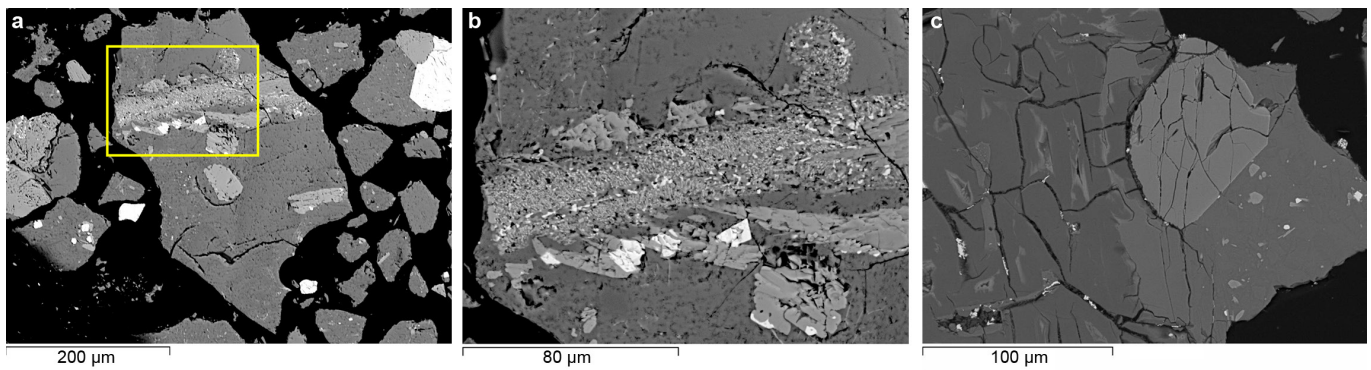
Extended Data Figure 1 | Grain size distribution of three volcanic ash samples collected during November 2012. At this proximal sampling location (275 m from the vent), most of the ash recovered is below 200 μm

in size and the dominant grain size peaks at around 50 μm. The measurements were made using a laser diffraction particle size analyser from Coulter.



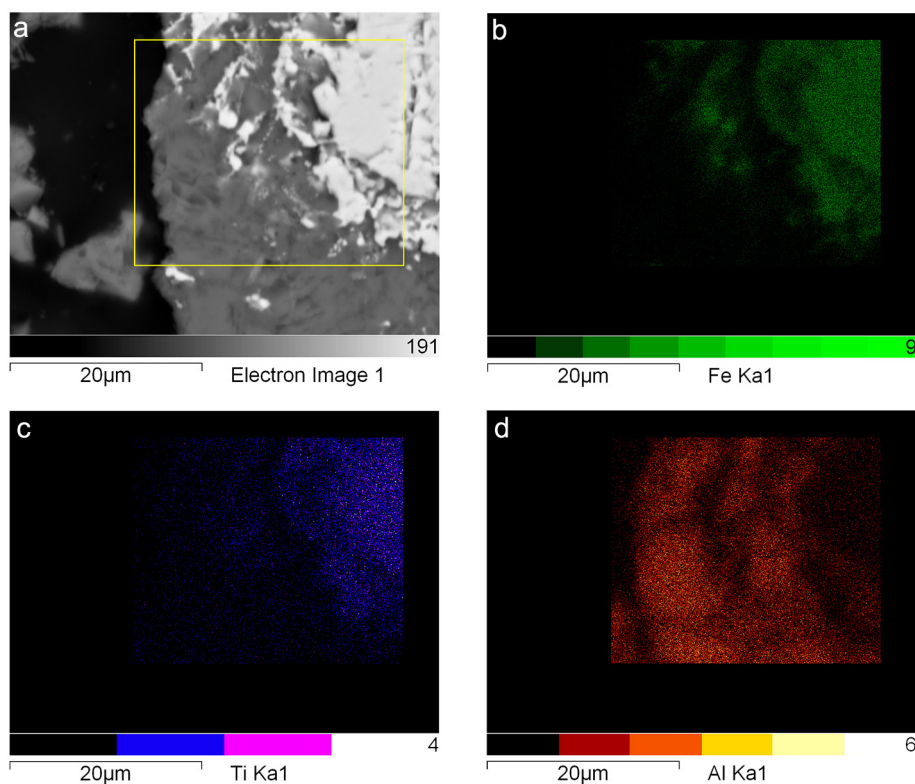
Extended Data Figure 2 | BSE image showing the different phases present in the eruptive products at Caliente. The dome rocks and the volcanic ash samples contain primarily plagioclase (Pl, dark grey), pyroxene (Px, light grey), iron oxides (Ox, white), apatite (Ap, very dark grey) and interstitial glass (Gl, dark grey). Note the absence of vesicles

(black, rounded pores) in this dense ash fragment, which contains less than 2% pore space. Despite the fact that there are no vesicles in this ash particle, we note that the edge of the iron oxides and pyroxene crystals are not straight, but rather crenulated and somewhat diffuse.



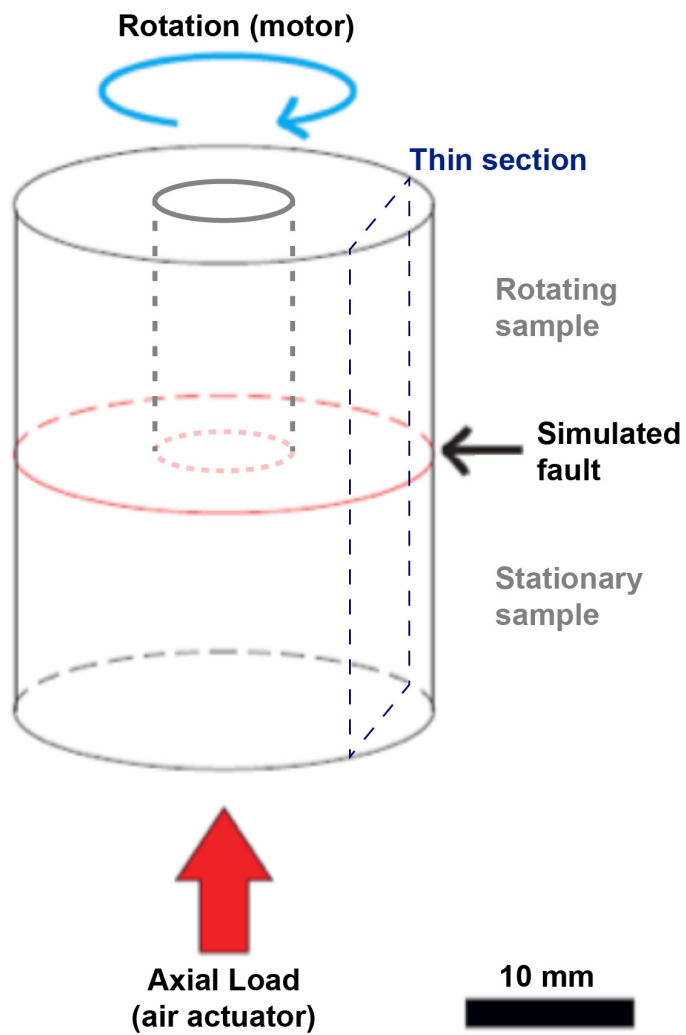
Extended Data Figure 3 | BSE images showing heterogeneous melt filaments present in volcanic ash erupted at Caliente. a, b, 13 November 2012; c, 26 November 2014. The yellow box in **a** defines the region of interest displayed in **b**. Evidence for high thermal input is best represented by the occurrence of frictional melting. The characteristic texture of

frictional melting has been noted in a number of volcanic ash particles and from several eruptions (the main text refers to ash from 10 November 2012). The textures associated with frictional melting preserved in the ash erupted on 13 November 2012 and 26 November 2014, suggest that this dynamic of strain localization in magma was active for at least two years.

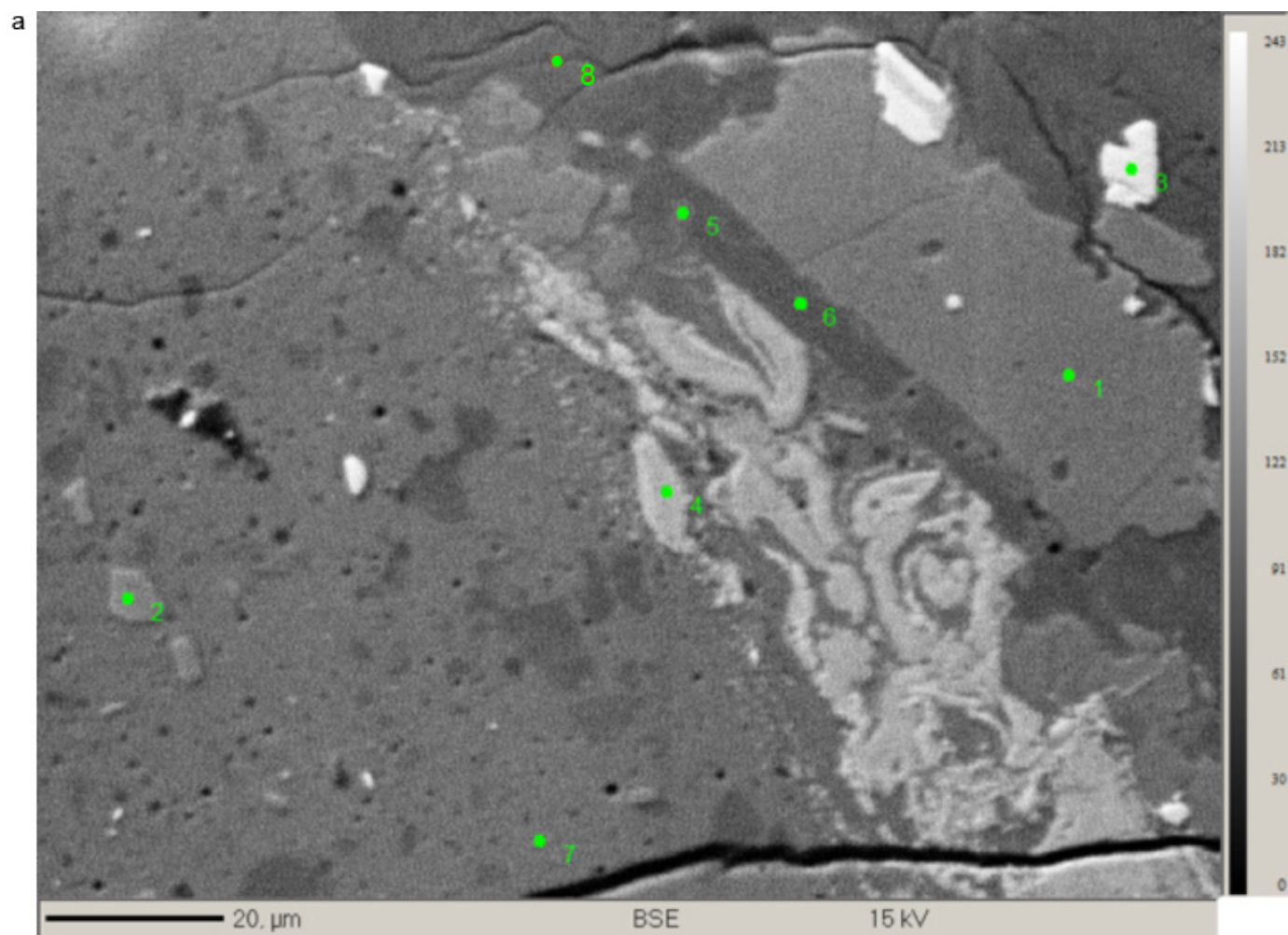


Extended Data Figure 4 | EDS images showing the heterogeneous concentration of various elements in the melt filaments. a, BSE image showing the area mapped by EDS. EDS maps show the distribution of Fe (**b**, in green), Ti (**c**, in blue), and Al (**d**, in red). Colour-scale values

represent X-ray counts per pixel for each energy band in the line type Ka1. During frictional melting of andesite and dacite, selective melting tends to affect the iron-titanium oxides more readily than silicate mineral phases owing to their lower fusion temperature³⁰.



Extended Data Figure 5 | Sample assembly setup during rotary shear experiments. The sketch also highlights the area sliced for thin-section preparation.

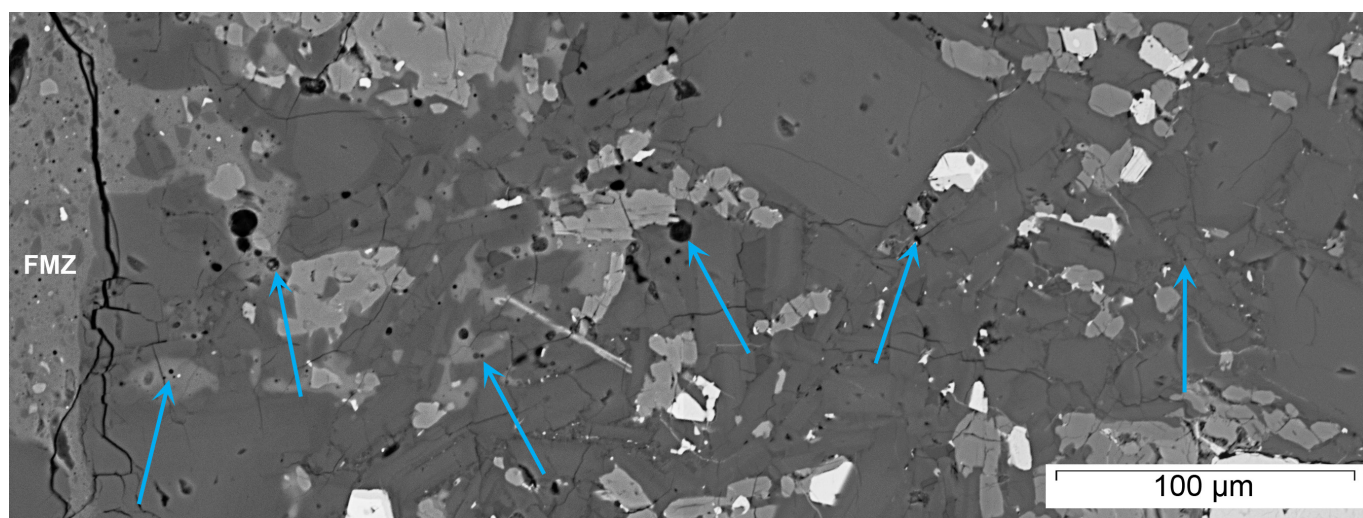


b

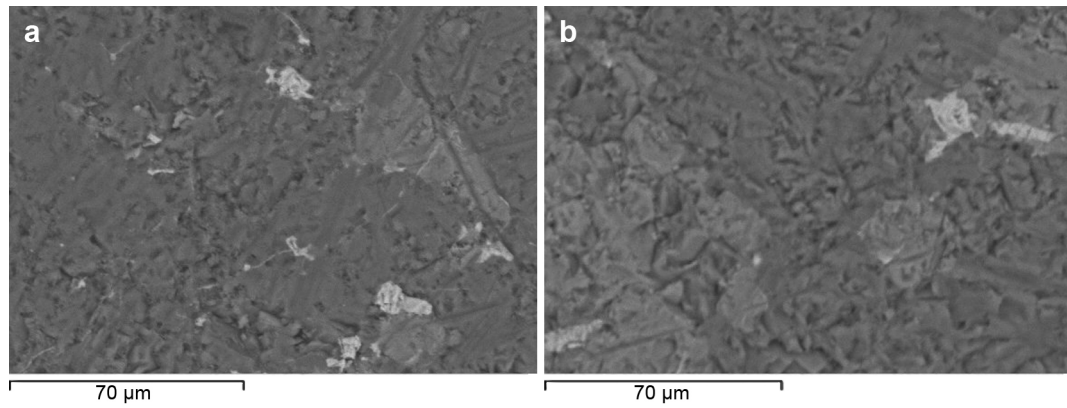
	#1	#2	#3	#4	#5	#6	#7	#8
Oxides	Clinopyroxene	Orthopyroxene	Fe-Oxide	Protomelt	Protomelt	Plagioclase	Frictional melt	Plagioclase
SiO ₂	51.17	52.62	1.30	24.50	54.91	53.70	53.28	56.59
Al ₂ O ₃	2.36	0.93	1.26	2.23	15.22	28.30	16.51	25.62
Na ₂ O	0.36	0.14	0.00	0.67	2.90	4.81	3.47	5.48
K ₂ O	0.02	0.03	0.17	0.40	0.69	0.25	0.94	0.73
MgO	15.32	18.06	0.59	7.75	7.98	0.09	6.76	0.64
CaO	16.66	6.32	0.31	0.85	8.37	11.43	7.37	9.15
TiO ₂	0.98	0.43	9.43	0.18	0.40	0.04	1.00	0.04
FeO	12.55	20.75	-	-	9.07	1.37	10.27	1.66
Fe ₂ O ₃	-	-	86.58	63.08	-	-	-	-
MnO	0.54	0.70	0.34	0.22	0.23	0.00	0.16	0.02
P ₂ O ₅	0.04	0.00	0.02	0.12	0.23	0.01	0.23	0.06
Total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Extended Data Figure 6 | Frictional melt chemistry. **a**, BSE image of the different phases and textures observed in the products of the rotary shear experiments, along with eight numbered locations of geochemical analyses acquired with the EPMA. **b**, Normalized geochemical composition of major elements for each analysis. Comparison of the chemical analyses with the textures reveals the variable heterogeneity of the rock products by frictional melting. Analyses 1 and 2 present pyroxene crystals in the seemingly undisturbed host rock and as fragments in the melt zone respectively; they do not show any degree of contamination. Similarly, analysis 3 presents a Fe-oxide crystal and analyses 6 and 8 present

plagioclase crystals in the host rock which have not been chemically altered by the products of frictional melting. Analysis 4 presents a protomelt consisting of orthopyroxene with high concentration of Fe-oxide. Analysis 5 also presents a protomelt but this time the chemical composition, and in particular the intermediate concentrations of MgO, CaO and FeO, suggests that it is a mixing product of molten plagioclase and orthopyroxene crystals in a ratio nearing 1:1. Analysis 7 presents the geochemistry of the more homogenized central frictional melt zone, resulting from the mixing of the molten crystals described above.



Extended Data Figure 7 | Vesicularity gradients developed in the interstitial glass along the edge of the FMZ. Blue arrows indicate vesicles. We observe no vesicles in the interstitial glass away (>0.4 mm) from the slip zone, and hence no vesicles in the pre-experimental sample.



Extended Data Figure 8 | SEM images showing the texture of dome rocks unchanged by subjecting them to 850°C. a, Heat applied for 30 min. **b,** Heat applied for 15 h. In either case, we note no new, spherical vesicles developed in the interstitial glass. This observation is consistent with the fact that the sample density did not change, as determined by

helium pycnometry. This observation indicates that even at atmospheric pressure, water is unable to exsolve at magmatic temperature, suggesting that high heat input is necessary to lower the solubility and increase diffusivity to trigger vesiculation.

Neonicotinoid pesticide exposure impairs crop pollination services provided by bumblebees

Dara A. Stanley¹, Michael P. D. Garratt², Jennifer B. Wickens², Victoria J. Wickens², Simon G. Potts² & Nigel E. Raine^{1,3}

Recent concern over global pollinator declines has led to considerable research on the effects of pesticides on bees^{1–5}. Although pesticides are typically not encountered at lethal levels in the field, there is growing evidence indicating that exposure to field-realistic levels can have sublethal effects on bees, affecting their foraging behaviour^{1,6,7}, homing ability^{8,9} and reproductive success^{2,5}. Bees are essential for the pollination of a wide variety of crops and the majority of wild flowering plants^{10–12}, but until now research on pesticide effects has been limited to direct effects on bees themselves and not on the pollination services they provide. Here we show the first evidence to our knowledge that pesticide exposure can reduce the pollination services bumblebees deliver to apples, a crop of global economic importance. Bumblebee colonies exposed to a neonicotinoid pesticide provided lower visitation rates to apple trees and collected pollen less often. Most importantly, these pesticide-exposed colonies produced apples containing fewer seeds, demonstrating a reduced delivery of pollination services. Our results also indicate that reduced pollination service delivery is not due to pesticide-induced changes in individual bee behaviour, but most likely due to effects at the colony level. These findings show that pesticide exposure can impair the ability of bees to provide pollination services, with important implications for both the sustained delivery of stable crop yields and the functioning of natural ecosystems.

Biotic pollination is required by a large proportion of crops worldwide¹⁰, disproportionately including those with economically high values and nutritional content¹³. The contribution of pollination services to global agriculture has been steadily increasing and was estimated at US\$361 billion in 2009 (ref. 14). In addition, animal-vector pollination is required by an estimated 87.5% of all angiosperms to reproduce¹¹, making this process fundamental to the functioning of natural ecosystems. Therefore, any threats to the delivery of pollination services could have serious consequences for both food security and wider ecosystem function. Neonicotinoid pesticides, the most widely used group of insecticides worldwide¹⁵, are implicated as one of the contributing factors in the global declines of bee pollinators^{3,16}. Although previous work has shown that bumblebee foraging activity, colony growth and reproduction can be altered by sublethal exposure to neonicotinoid pesticides^{1,2,5–7}, all research on pesticide effects has focused on bees as the service providers, but has not assessed the pollination service itself. Therefore it is unknown whether pesticide exposure actually results in changes to the delivery of pollination services to crops and wild plants (for a discussion of potential mechanisms see ref. 17). This information is essential to assess the severity of pesticide effects on ecosystem services, and to inform actions to mitigate negative effects.

Apples are an important global crop, with 75 million tonnes harvested from 95 countries in 2012 and an estimated export value of US\$71 billion (Food and Agriculture Organisation statistics, <http://faostat3.fao.org>). Apple crops benefit from insect pollination with seed number, fruit set, fruit size and shape all improved with increased

pollination services¹⁸. Bumblebees are major pollinators of apples¹⁹ and many other crops across the world¹², and are exposed to low levels of pesticides when foraging in agricultural areas. Here we investigated how exposure to low, field-realistic levels of a widely used neonicotinoid insecticide (thiamethoxam) could affect the ability of bumblebees to pollinate apple trees. We pre-exposed colonies to 2.4 parts per billion (ppb) thiamethoxam, 10 ppb thiamethoxam or control solutions (containing no pesticide; rationale for selecting pesticide concentrations and relevance of results are outlined in Methods and Supplementary Information) in their nectar source (artificial sugar water) for a period of 13 days (8 colonies per treatment, that is, 24 colonies in total). Subsequently, colonies were brought to the field and allowed access to virgin apple trees of a dessert (Scrumptious) variety, along with trees of a polliniser (Everest) variety, in pollinator exclusion cages in which we observed both individual- and colony-level behaviour. At the end of the season, apples from tested trees were collected to assess pollination service delivery in terms of fruit and seed set.

When whole colonies were given access to apple trees we found an effect of insecticide treatment on visitation rates to apple flowers ($F_{2,86} = 3.1$, $P = 0.05$); colonies exposed to 10 ppb pesticide provided lower visitation rates to apple flowers than controls (Fig. 1a; Extended Data Table 1). We also found an effect of treatment on the number of foraging trips from which bees returned carrying pollen ($\chi^2 = 9.65$, degrees of freedom (df) = 2, $P = 0.008$), with fewer bees from colonies exposed to 10 ppb pesticide returning with pollen than workers from control colonies (Fig. 1b). Apple abortion rate was affected by treatment ($\chi^2 = 5.94$, df = 2, $P = 0.05$), with trees pollinated by

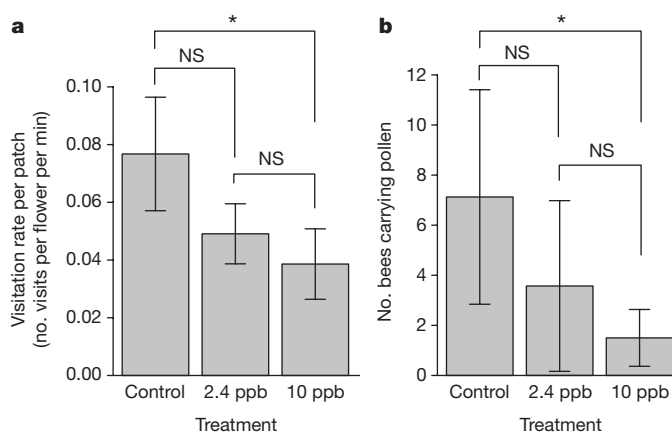


Figure 1 | Effects of pesticide treatment on colony-level behaviour.

a, b, Visitation rates provided by colonies to Scrumptious apple flowers (number of visits per flower per minute) (**a**) and number of foraging trips from which bees returned carrying pollen (**b**), from colonies exposed to different pesticide treatments. Eight colonies were observed per treatment group, and means \pm s.e.m. are shown, * $P < 0.05$. NS, not significant. Results from statistical models are given in Extended Data Table 1.

¹School of Biological Sciences, Royal Holloway University of London, Egham TW20 0EX, UK. ²Centre for Agri-Environmental Research, School of Agriculture, Policy and Development, University of Reading, Reading RG6 6AR, UK. ³School of Environmental Sciences, University of Guelph, Guelph, Ontario N1G 2W1, Canada.

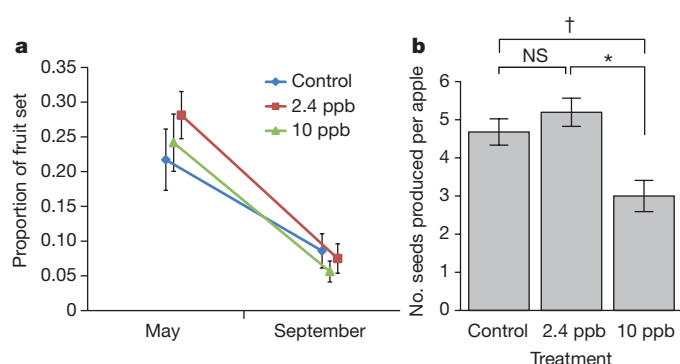


Figure 2 | Effects of pesticide treatment on fruit and seed set.

a, b, The change in proportion of fruit set for trees (48 trees in total, 16 per treatment) pollinated by colonies exposed to different pesticide treatments measured early (May) and late (September), which represents fruit abortion level (**a**), and number of seeds produced per apple (134 apples in total; 53 in control, 46 in 2.4 ppb and 35 in 10 ppb pesticide treatments) pollinated by colonies exposed to different pesticide treatments (**b**). Eight colonies were observed per treatment group, and means \pm s.e.m. are shown, * $P < 0.05$, † indicates a difference of $P = 0.06$ between control and 10 ppb. NS, not significant. Results from statistical models are given in Extended Data Table 1.

2.4 ppb pesticide-exposed colonies aborting more fruit than controls (Fig. 2a), although overall levels of fruit set did not differ ($\chi^2 = 4.1$, $df = 2$, $P = 0.13$) and there was no difference in the proportion of trees that produced fruit among treatments ($\chi^2 = 1.2$, $df = 2$, $P = 0.55$). However, we found a significant effect of treatment on the number of seeds produced per apple, an indicator of fruit quality, ($\chi^2 = 8.27$, $df = 2$, $P = 0.02$); flowers pollinated by colonies exposed to 10 ppb pesticide produced significantly fewer seeds than those pollinated by 2.4 ppb colonies (Fig. 2b). These results show that colonies exposed to pesticide can deliver reduced pollination services to apple crops.

These colony-level effects could be explained by several mechanisms, including individual behavioural changes. Individual bees exposed to 10 ppb pesticide spent longer foraging ($F_{2,57} = 3.72$, $P = 0.03$; Fig. 3a), visited more *Scrumptious* flowers ($\chi^2 = 12.79$, $df = 2$, $P = 0.002$) and switched more frequently between varieties during each trip ($\chi^2 = 11.32$, $df = 2$, $P = 0.003$; Fig. 3b; Extended Data Table 2), which suggests a modification of their floral preferences⁷. Neonicotinoids target neurotransmitter receptors in insects and, as well as causing neuronal inactivation²⁰, some have been shown to be

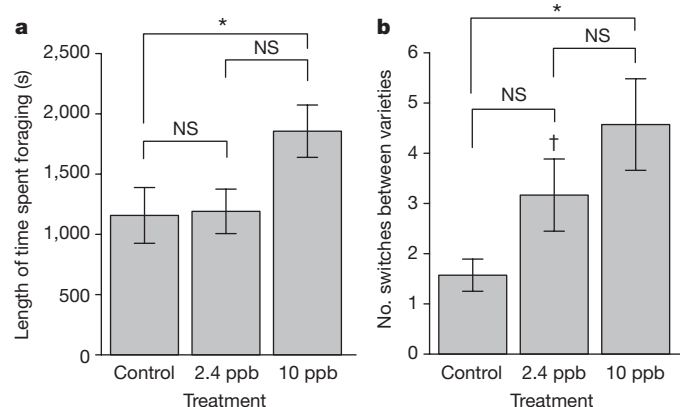


Figure 3 | Effects of pesticide treatment on individual bee behaviour.

a, b, Time spent foraging per foraging trip (seconds; $n = 68$ bees) (**a**) and number of switches between *Scrumptious* and *Everest* apple varieties ($n = 93$ bees) (**b**) for individual bees exposed to different pesticide treatments. Means \pm s.e.m. are shown, * $P < 0.05$, † indicates a difference of $P = 0.06$ between control and 2.4 ppb. NS, not significant. Results from statistical models are given in Extended Data Table 2.

partial neuronal agonists²¹; therefore increases in individual foraging activity may be explained by acute increases in neuronal activity causing hormesis (a biphasic response in which low levels of an otherwise toxic compound can result in stimulation of a biological process²²). However, we found no effect of treatment on whether flowers visited by these individual bees produced apples ($\chi^2 = 0.88$, $df = 2$, $P = 0.64$), showed higher rates of fruit abortion ($\chi^2 = 0.42$, $df = 2$, $P = 0.81$) or different levels of seed set ($\chi^2 = 0.11$, $df = 2$, $P = 0.95$). This suggests that bees exposed to pesticide must somehow be behaving differently on flowers, in a way that was not readily observable in our experiment (for example, changes in stigmatic contact²³), such that increased visit frequency did not result in better pollination service delivery at the individual level.

Our results suggest that effects on pollination service delivery are not due to individual behavioural modification, but instead are most likely due to changes in colony activity levels as evidenced by reduced floral visitation rates and pollen collection. Bees collecting pollen may be more effective pollinators as they can deposit more pollen on plant stigmas²⁴; therefore if pesticide-exposed colonies are collecting less pollen they are also likely to be depositing less on stigmas than bees from control colonies. While individual bees exposed to pesticides visited more flowers, overall pesticide-exposed colonies provided lower visitation rates and collected less pollen, thus explaining why reduced pollination services were delivered. Gill & Raine⁷ found that control (untreated) bees improved their pollen foraging performance over time, whereas imidacloprid-treated bees became less successful foragers; foragers in our colony-level experiment may have carried out multiple trips and become more experienced foragers, potentially explaining why we find effects on pollen collection here but not in the individual-level experiment. Interestingly, for almost all parameters measured in this study we found significant effects on both individual behaviour and colony-level function following 10 ppb thiamethoxam exposure, but not at the 2.4 ppb level. This suggests that there are dose-dependent effects that lie between these two exposure levels. Both these exposure levels are highly relevant as they are within the range measured in the field, but further work is necessary to elucidate the lowest level at which these effects become significant (for further discussion of rationale for exposure and relevance of results, see Methods and Supplementary Information).

A 36% reduction in the number of seeds produced in apples pollinated by colonies exposed to 10 ppb pesticide in comparison to control colonies has important agronomic implications for crop production. The number of seeds in apples is closely linked to fruit crop quality in most, but not all, varieties^{18,25} and the enhancement of fruit quality, particularly the proportion of Class 1 fruit, underpins the economic value of UK orchards²⁶; growers must typically thin out their apple crops making the quality of each fruit very important. Therefore impacts on seed set and fruit quality have direct implications for apple production value, and as seed set and fruit set are positively linked in many varieties, reduced seed set can have direct negative implications for fruit set and total crop yield^{26,27}. As certain apple varieties in the UK currently experience pollination deficits^{19,26}, mitigating the effects of pesticides on bumblebee pollinators could improve pollination service delivery. Apple crops are visited by a wide variety of pollinator groups, and neonicotinoid pesticides differentially affect insect taxa^{4,28}. Apart from bumblebees, one of the other main pollinator groups that visit apple flowers are solitary bees¹⁹, and it has been suggested that pesticide sensitivity of solitary bees is likely to be higher than for larger, social species like bumblebees^{4,5,17,29}. Therefore, apple pollination in a field setting could be more vulnerable to pesticide exposure than measured here.

Bumblebees are essential pollinators of many important crops other than apples, including field beans, berries, tomatoes and oilseed rape^{12,26}. If exposure to pesticides alters pollination services to apple crops, it is likely that these other bee-pollinated crops would also be affected. Most importantly, the majority of wild plant species benefit from insect

pollination services¹¹. Therefore reduced pollination by pesticide-affected colonies, as evidenced by reduced seed set, also has significant implications for pollination in wild systems. Many wild plant species are both self-incompatible and pollen limited³⁰, so any reduction in the delivery of pollination services could have substantial effects on wild plant communities and therefore wider ecosystem function.

Concerns over global bee declines are strongly driven by the need for the essential pollination services they provide to both crops and wild plants. The use of neonicotinoid pesticides presents a potential threat to bee health and, although the evidence base reporting sublethal (behavioural) effects of pesticides on bees is mounting³, we have shown for the first time that there is also an important effect of pesticide exposure on the pollination services bees provide. This information provides a new perspective when trying to fully understand the trade-offs involved when using insecticides, showing that both the potential benefits and the true costs of pest control options need to be considered.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 July; accepted 23 October 2015.

Published online 18 November 2015.

- Gill, R. J., Ramos-Rodriguez, O. & Raine, N. E. Combined pesticide exposure severely affects individual- and colony-level traits in bees. *Nature* **491**, 105–108 (2012).
- Whitehorn, P. R., O'Connor, S., Wäckers, F. L. & Goulson, D. Neonicotinoid pesticide reduces bumble bee colony growth and queen production. *Science* **336**, 351–352 (2012).
- Godfray, H. C. J. *et al.* A restatement of the natural science evidence base concerning neonicotinoid insecticides and insect pollinators. *Proc. R. Soc. Lond. B* **281**, 20140558 (2014).
- Arena, M. & Sgolastra, F. A meta-analysis comparing the sensitivity of bees to pesticides. *Ecotoxicology* **23**, 324–334 (2014).
- Rundlöf, M. *et al.* Seed coating with a neonicotinoid insecticide negatively affects wild bees. *Nature* **521**, 77–80 (2015).
- Feltham, H., Park, K. & Goulson, D. Field realistic doses of pesticide imidacloprid reduce bumblebee pollen foraging efficiency. *Ecotoxicology* **23**, 317–323 (2014).
- Gill, R. J. & Raine, N. E. Chronic impairment of bumblebee natural foraging behaviour induced by sublethal pesticide exposure. *Funct. Ecol.* **28**, 1459–1471 (2014).
- Henry, M. *et al.* A common pesticide decreases foraging success and survival in honey bees. *Science* **336**, 348–350 (2012).
- Fischer, J. *et al.* Neonicotinoids interfere with specific components of navigation in honeybees. *PLoS One* **9**, e91364 (2014).
- Klein, A. M. *et al.* Importance of pollinators in changing landscapes for world crops. *Proc. R. Soc. Lond. B* **274**, 303–313 (2007).
- Ollerton, J., Winfree, R. & Tarrant, S. How many flowering plants are pollinated by animals? *Oikos* **120**, 321–326 (2011).
- Kleijn, D. *et al.* Delivery of crop pollination services is an insufficient argument for wild pollinator conservation. *Nat. Commun.* **6**, 7414 (2015).
- Eilers, E. J., Kremen, C., Greenleaf, S. S., Garber, A. K. & Klein, A. M. Contribution of pollinator-mediated crops to nutrients in the human food supply. *PLoS One* **6**, e21363 (2011).
- Lautenbach, S., Seppelt, R., Liebscher, J. & Dormann, C. F. Spatial and temporal trends of global pollination benefit. *PLoS One* **7**, e35954 (2012).
- Goulson, D. An overview of the environmental risks posed by neonicotinoid insecticides. *J. Appl. Ecol.* **50**, 977–987 (2013).

- Goulson, D., Nicholls, E., Botías, C. & Rotheray, E. L. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science* **347**, 1255957 (2015).
- Brittain, C. & Potts, S. G. The potential impacts of insecticides on the life-history traits of bees and the consequences for pollination. *Basic Appl. Ecol.* **12**, 321–331 (2011).
- Sheffield, C. S. Pollination, seed set and fruit quality in apple: studies with *Osmia lignaria* (Hymenoptera: Megachilidae) in the Annapolis Valley, Nova Scotia, Canada. *J. Pollinat. Ecol.* **12**, 120–128 (2014).
- Garratt, M. P. D. *et al.* Pollination deficits in UK apple orchards. *J. Pollinat. Ecol.* **12**, 9–14 (2014).
- Palmer, M. J. *et al.* Cholinergic pesticides cause mushroom body neuronal inactivation in honeybees. *Nat. Commun.* **4**, 1634 (2013).
- Dégise, P., Grünwald, B. & Gauthier, M. The insecticide imidacloprid is a partial agonist of the nicotinic receptor of honeybee Kenyon cells. *Neurosci. Lett.* **321**, 13–16 (2002).
- Cutler, G. C. & Rix, R. R. Can poisons stimulate bees? Appreciating the potential of hormesis in bee-pesticide research. *Pest Manag. Sci.* **71**, 1368–1370 (2015).
- Sakamoto, R. L., Morinaga, S.-I., Ito, M. & Kawakubo, N. Fine-scale flower-visiting behavior revealed by using a high-speed camera. *Behav. Ecol. Sociobiol.* **66**, 669–674 (2012).
- Castro, S., Loureiro, J., Ferrero, V., Silveira, P. & Navarro, L. So many visitors and so few pollinators: variation in insect frequency and effectiveness governs the reproductive success of an endemic milkwort. *Plant Ecol.* **214**, 1233–1245 (2013).
- Buccheri, M. & Di Vaio, C. Relationship among seed number, quality, and calcium content in apple fruits. *J. Plant Nutr.* **27**, 1735–1746 (2004).
- Garratt, M. P. D. *et al.* Avoiding a bad apple: Insect pollination enhances fruit quality and economic value. *Agric. Ecosyst. Environ.* **184**, 34–40 (2014).
- Volz, R. K., Tustin, D. S. & Ferguson, I. B. Pollination effects on fruit mineral composition, seeds and cropping characteristics of 'Braeburn' apple trees. *Sci. Hortic. (Amsterdam)* **66**, 169–180 (1996).
- Biddinger, D. J. *et al.* Comparative toxicities and synergism of apple orchard pesticides to *Apis mellifera* (L.) and *Osmia cornifrons* (Radoszkowski). *PLoS One* **8**, e72587 (2013).
- Decourtye, A., Henry, M. & Desneux, N. Environment: overhaul pesticide testing on bees. *Nature* **497**, 188 (2013).
- Burd, M. Bateman's principle and plant reproduction: the role of pollen limitation in fruit and seed set. *Bot. Rev.* **60**, 83–139 (1994).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank technicians at the University of Reading for assistance in apple collection and seed counting, and E. van Leeuwen and colleagues at Royal Holloway University of London for useful discussions. This study was supported by UK Insect Pollinators Initiative grants BB/I000178/1 awarded to N.E.R. and BB/1000348/1 awarded to S.G.P. (funded jointly by the Living with Environmental Change programme, Biotechnology and Biological Sciences Research Council (BBSRC), Wellcome Trust, Scottish Government, Department for Environment, Food and Rural Affairs (Defra) and Natural Environment Research Council (NERC)). N.E.R. is supported as the Rebanks Family Chair in Pollinator Conservation by The W. Garfield Weston Foundation.

Author Contributions D.A.S. and N.E.R. conceived the project, D.A.S., N.E.R. and M.P.D.G. designed the research, D.A.S., J.B.W. and V.J.W. carried out the research, D.A.S., N.E.R., M.P.D.G. and S.G.P. contributed equipment for the research, D.A.S. analysed the data, all authors were involved in writing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.A.S. (darastanley@gmail.com) or N.E.R. (nraine@uoguelph.ca).

METHODS

Pesticide preparation. A stock pesticide solution was made by dissolving 100 mg thiamethoxam (PESTANAL, Analytical Standard, Sigma Aldrich) in 100 ml acetone (1 mg ml^{-1}). Aliquots of stock solution were added to 40% sucrose to create treatment solutions of $10 \mu\text{g l}^{-1}$ (10 ppb) and $2.4 \mu\text{g l}^{-1}$ (2.4 ppb) thiamethoxam. These concentrations were chosen as field-realistic; the lower concentration (2.4 ppb) was based on thiamethoxam concentrations found in nectar pots of bumblebee colonies foraging in agricultural areas in the UK³¹ and in pollen collected by honeybees³², and the higher concentration (10 ppb) is within the range measured in pollen and nectar and of a variety of treated crops^{33–35} and contaminated wild flowers^{35–37}, and has been used in previous studies examining effects of another neonicotinoid (imidacloprid) on bumblebee behaviour^{1,7}. A control solution was also made by repeating the process outlined above but using an aliquot of 10 ppb acetone only (that is, no pesticide).

Experimental setup. Twenty-four commercially reared *Bombus terrestris audax* colonies were obtained from Biobest (Westerlo, Belgium) at the start of April 2014, each containing a queen and an average of 99 workers (range 57–133). Colonies were weighed on arrival to estimate the overall colony size, and each assigned sequentially to one of three treatment groups (2.4 ppb thiamethoxam, 10 ppb thiamethoxam and control) based on decreasing mass (but randomly assigned within block). Each day, three colonies (one from each treatment) were assigned to treatment groups, until after 7 days all colonies were receiving treated sucrose (16 colonies exposed to thiamethoxam and 8 to control solution). We chose this sequential exposure regime to mimic subsequent field testing and ensure all colonies had comparable durations of exposure to their treatment. Colonies were fed treated sucrose solution from a gravity feeder inserted at the base of the nest box. Feeders were initially refilled every 2–3 days, and then every 1–2 days when the colonies had grown significantly. Untreated, defrosted honeybee-collected pollen was provided to colonies every 2–3 days. Colonies were exposed to treatments for an average of 13 days (range 12–15) before field testing. Before being moved to the field, colonies had access to a feeder containing sucrose (40%) in a laboratory flight arena for 48 h to become accustomed to leaving the nest to forage. There was no difference in colony weights at the start (ANOVA: $F_{2,21} = 0.091$, $P = 0.91$) or end (ANOVA: $F_{2,21} = 0.88$, $P = 0.43$) of the experimental period, indicating no treatment effect on colony size.

Field testing. Cage experiments were carried out at Sonning Farm, University of Reading, UK. 100 apple trees of a commercial dessert apple (Scrumptious variety) were moved into holding pollinator exclusion cages in mid-March 2014 before flowering to prevent insect visitation. Field experiments began when trees were entering full flower in mid-April. Each day, one colony from each treatment was taken from the laboratory, placed individually in one of the three test cages and observed simultaneously (with one observer per cage) in a randomized block design (see below for details of observations). Each day a different treatment was assigned to each observer. Cages were $4.8 \times 2.1 \times 2.1 \text{ m}$ frames covered in polyethylene mesh (gauge size = 1.33 mm, Extended Data Fig. 1). Observations were carried out on 8 dry, bright days from 16–26 April 2014 spanning the peak flowering of apples (daily means: maximum temperature 16°C , rainfall 2.5 mm). This flowering period limited the number of days on which testing could be carried out, and therefore the number of colonies that could be tested; as a result no statistical methods were used to predetermine sample size. The investigators were not blinded to allocation during experiments and outcome assessment.

Individual-level measurements. Each morning, three cages were populated with two virgin Scrumptious trees each from the holding cages (mean \pm s.e.m. = 130 ± 8.5 flowers per tree) as well as two polliniser trees (Everest variety, mean \pm s.e.m. = 305 ± 15 flowers per tree, Extended Data Fig. 1). The number of flowers of each variety was standardized across cages to ensure equal floral density each day, and 40 open and receptive flowers were marked with cable ties on each Scrumptious tree for subsequent estimation of pollination services (fewer flowers were marked on the last day of observations as there were no longer 40 full-bloom flowers—flower numbers on these days were noted). The nest boxes in each cage were then opened to allow a single worker to exit. This bee was observed for the duration of its foraging trip (until it attempted to return to the nest), or until 60 min had elapsed (Extended Data Fig. 2). The duration of the foraging trip, the number of flowers of each apple variety visited, and the handling time for each flower visit was recorded using Etholog software (EthoLog: Behavioural observation transcription tool, University of Sao Paulo, Brazil, 2011). If the individual bee did not visit any flowers within the first 20 min, it was assumed not to be a forager and was captured, returned to the colony and another bee released. All bees that foraged were paint-marked before they were returned to the colony to ensure the same individuals were not observed twice.

This process was repeated until all cages had the same number of active foragers recorded (3–5 bees per colony each day). Individual level observations took place between 10:00 and 16:30.

Colony-level measurements. After individual-level observations, the two focal Scrumptious trees in each cage were removed and replaced with two new virgin trees. Again we standardized the number of flowers of each variety across cages with 40 open and receptive flowers on each tree marked with cable ties. Colony boxes were opened to allow free entry and exit to all active bees for a period of 60 min. This time period was chosen to avoid over-pollination of test flowers based on pilot observations. Colony activity was monitored at the nest entrance using video cameras. After an initial 10-min period to allow the bees to become accustomed to the setup, four 10-min focal observations were carried out on separate patches of Scrumptious flowers in each cage to estimate visitation rates. At the end of the 60-min period, the Scrumptious trees were immediately removed to prevent further visitation. Colony level observations were carried out between 14:30 and 18:30.

Estimation of pollination services. At the end of both the individual and colony observation periods, all test trees were returned to holding cages in which they were not visited by any other insects until apples were harvested at the end of the season. An initial assessment of fruit set from marked flowers (indicating flowers open during cage tests) was made at the end of May for all test Scrumptious trees to assess how many flowers were proceeding to fruit set stage (and how many aborted, Fig. 2a). Marked apples were collected on 27 August, and a final assessment made of the proportion of marked flowers that had produced mature fruit (Extended Data Fig. 2). In the lab, seed number was counted per apple for all collected fruit (274 apples from 96 trees across both experiments). Details of all data analyses carried out are given in the supplementary information.

Data analysis. *Individual level.* Measures of the number of flowers visited, numbers of switches between apple varieties, duration of total time in cage (from when the bee left the colony box until it returned/end of 60 min period) and time taken to visit the first flower (latency) were recorded for all individual bees. For 68 of 93 bees observed (evenly distributed across cages and treatments) a number of additional response variables were also recorded including mean duration of the first 5 flower visits, number of inter flower intervals longer than 60 s, mean duration of flower visits, mean period of time between flower visits, length of time spent foraging (time between first and last flower visit) and total time spent on flowers (sum of durations for all individual flower visits). We tested for differences in these measures among treatments by constructing mixed-effects models with pesticide treatment as a fixed effect. As several variables differed among days, including weather, floral abundance and the identity of colonies used, day of testing was included as a random blocking factor in all models. Data were analysed in R version 3.1.0 (ref. 38), using either linear mixed effects (LME) models with the lmer function in the nlme package for continuous data³⁹, generalized mixed effects (GLMM) models with Poisson distribution used for response variables that were counts using the glmer function in the lme4 package⁴⁰, or the glmmPQL function in the MASS package⁴¹ when data were overdispersed. Models were validated by plotting standardized residuals versus fitted values, normal qq-plots and histograms of residuals, and continuous response variables were logarithmically transformed ($\log(X + 1)$) if necessary to improve residual fit. If treatment was significant, Tukey's post hoc tests were performed using the glht function in the multcomp package⁴².

To assess differences in apple production on trees visited by pesticide exposed and control bees, we examined a number of variables including the number of fruits produced at the start of the season (May) and at the end (September; Fig. 2a), the change in proportion of apples forming from marked flowers per tree between the start and end measures (fruit abortion levels) and number of seeds per apple (measured in early September; Fig. 2b). Models were run as described previously with treatment as a fixed effect, although the tree on which fruits were produced, the number of bees released and date of testing were included as random effects. As a number of trees produced no fruit, seed set data were analysed in two steps. First, we tested whether there was a treatment difference in the number of trees that produced any fruit. Second, we tested for treatment differences in seeds per apple (a measure that only included trees that had produced some fruit).

Colony level. We tested for differences in colony activity levels (the combined number of entries and exits by workers to the colony box) and the number of bees carrying pollen among treatments using GLMM models in the MASS package⁴¹, with Poisson distribution for count data. Treatment differences in flower visitation rate to Scrumptious trees were tested using LME models³⁹. Date of testing was used as a random effect in all models (and patch included as a random effect in the flower visitation rate model), and models were validated as

described above. Fruit abortion and seed set variables were analysed as described for the individual level experiment, using tree and date of testing as random effects.

31. Thompson, H. *et al.* Effects of neonicotinoid seed treatments on bumble bee colonies under field conditions. (Food and Environment Research Agency (FERA), 2013).
32. Pilling, E., Campbell, P., Coulson, M., Ruddle, N. & Tornier, I. A four-year field program investigating long-term effects of repeated exposure of honey bee colonies to flowering crops treated with thiamethoxam. *PLoS One* **8**, e77193 (2013).
33. Castle, S. J., Byrne, F. J., Bi, J. L. & Toscano, N. C. Spatial and temporal distribution of imidacloprid and thiamethoxam in citrus and impact on *Homalodisca coagulata* populations. *Pest Manag. Sci.* **61**, 75–84 (2005).
34. Dively, G. P. & Kamel, A. Insecticide residues in pollen and nectar of a cucurbit crop and their potential exposure to pollinators. *J. Agric. Food Chem.* **60**, 4449–4456 (2012).
35. Botías, C. *et al.* Neonicotinoid residues in wildflowers, a potential route of chronic exposure for bees. *Environ. Sci. Technol.* **9**, 12731–12740 (2015).
36. Krupke, C. H., Hunt, G. J., Eitzer, B. D., Andino, G. & Given, K. Multiple routes of pesticide exposure for honey bees living near agricultural fields. *PLoS One* **7**, e29268 (2012).
37. Stewart, S. D. *et al.* Potential exposure of pollinators to neonicotinoid insecticides from the use of insecticide seed treatments in the mid-southern United States. *Environ. Sci. Technol.* **48**, 9762–9769 (2014).
38. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org> (2011).
39. Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & R Development Core Team. Package “nlme”: Linear and nonlinear mixed effects models. R package version 3.1-104 (2012).
40. Bates, D., Maechler, M., Bolker, B. & Walker, S. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7 <http://CRAN.R-project.org/package=lme4> (2014).
41. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S*. 4th edn (Springer, 2002).
42. Hothorn, T., Bretz, F. & Westfall, P. Simultaneous inference in general parametric models. *Biom. J.* **50**, 346–363 (2008).



Extended Data Figure 1 | An example of the experimental setup at the Sonning Farm field site. Experimental pollinator exclusion cages containing a bumblebee colony (located in the corner of the cage) and potted experimental apple trees are shown. Photos: D.A.S.



Extended Data Figure 2 | An experimental bumblebee (*Bombus terrestris*) worker visiting an apple flower (left), and an example of an apple produced from a marked (yellow cable tie) apple flower (right; *Scrumptious* variety). Photos: D.A.S. and C. L. Truslove.

Extended Data Table 1 | Results from the colony-level experiment

Colony level	Mean \pm SE			Model summary			
<i>Activity</i>	control	2.4ppb	10ppb		df	p	
Total no. of entrances and exits to colony	53.9 \pm 22.3	44.3 \pm 12.5	25.3 \pm 12	$\chi^2=4.19$	2	0.12	glmmPQL
No. of bee visits returning with pollen	7.13 \pm 4.28	3.57 \pm 3.41	1.5 \pm 1.13	$\chi^2=9.65$	2	0.008	glmmPQL
Visitation rate to Scrumptious flowers (no flowers/bee/minute)	0.08 \pm 0.02	0.05 \pm 0.01	0.04 \pm 0.01	F=3.1	2,86	0.05	lme
<i>Fruit set</i>							
Start no. of fruit	8.63 \pm 1.78	11.25 \pm 1.35	9.68 \pm 1.65	$\chi^2=2.67$	2	0.26	glmmPQL
End no. of fruit	3.44 \pm 0.99	3 \pm 0.85	2.25 \pm 0.60	$\chi^2=4.1$	2	0.13	glmer
Change in proportion of fruit between May & Sept (Abortion rate)	0.13 \pm 0.29	0.21 \pm 0.03	0.19 \pm 0.03	$\chi^2=5.94$	2	0.05	glmmPQL
Proportion of trees producing apples	0.69	0.69	0.81	$\chi^2=1.2$	2	0.55	glmer
Seed no. per apple	4.68 \pm 0.3	5.2 \pm 0.4	3 \pm 0.4	$\chi^2=8.27$	2	0.02	glmer

Significant differences ($P \leq 0.05$) are highlighted in bold.

Extended Data Table 2 | Results from the individual-level experiment

Individual level	Mean \pm SE			Model summary			
<i>Behaviour</i>	control	2.4ppb	10ppb		df	p	
Latency to first flower visit (secs)	339 \pm 55	289 \pm 41	245 \pm 49	F=2.8	2,79	0.07	lme
Mean duration of first 5 flower visits (secs)	12 \pm 1.3	18 \pm 2.7	16 \pm 2.3	F=1.97	2,58	0.15	lme
No. interflower intervals longer than 60 secs	1.7 \pm 0.5	1.4 \pm 0.3	2.1 \pm 0.6	F=0.59	2,57	0.55	lme
Total no. flowers visited	83 \pm 15	97 \pm 14	125 \pm 22	χ^2 =4.65	2	0.1	glmmPQL
Total no. Everest flowers visited	55.4 \pm 15	49.6 \pm 12	71.3 \pm 21	χ^2 =1.9	2	0.39	glmmPQL
Total no. Scrumptious flowers visited	27.6 \pm 5.1	47.3 \pm 5.7	53.5 \pm 5.7	χ^2 =12.79	2	0.002	glmmPQL
Proportion of bees that collected pollen	0.38	0.22	0.19	χ^2 =2.63	2	0.27	glmer
Mean duration of flower visits (secs)	7 \pm 1	8 \pm 1.2	11 \pm 3.5	F=0.98	2,58	0.38	lme
Mean period of time between flower visits (secs)	28 \pm 10	11 \pm 3.1	17 \pm 5.1	F=0.76	2,58	0.47	lme
Length of time spent foraging (time of last flower visit - time of first flower visit) (secs)	1157 \pm 231	1191 \pm 184	1856 \pm 217	F=3.72	2,57	0.03	lme
Total length of time spent on flowers (secs)	375 \pm 55	539 \pm 66	762 \pm 113	F=7.35	2,57	0.001	lme
Duration of total time in cage (secs)	2041 \pm 239	2162 \pm 202	2383 \pm 204	F=1.338	2,84	0.27	lme
Total no. of switches between apple varieties	1.57 \pm 0.3	3.17 \pm 0.7	4.57 \pm 0.9	χ^2 =11.32	2	0.003	glmmPQL
<i>Fruit set</i>							
Start no. of fruit	8.13 \pm 1.28	9.50 \pm 1.94	9.69 \pm 2.03	χ^2 =0.53	2	0.77	glmer
End no. of fruit	1.38 \pm 0.52	2.06 \pm 0.80	3.13 \pm 1.17	χ^2 =3.82	2	0.15	glmer
Change in proportion of fruit between May & Sept (Abortion rate)	0.17 \pm 0.03	0.19 \pm 0.04	0.16 \pm 0.03	χ^2 =0.42	2	0.81	glmer
Proportion of trees producing apples	0.44	0.5	0.56	χ^2 =0.88	2	0.64	glmer
Seed no. per apple	3.52 \pm 0.6	3.66 \pm 0.4	3.02 \pm 0.4	χ^2 =0.11	2	0.95	glmmPQL

Significance differences ($P \leq 0.05$) are highlighted in bold.

Exceptional preservation of tiny embryos documents seed dormancy in early angiosperms

Else Marie Friis^{1,2}, Peter R. Crane², Kaj Raunsgaard Pedersen^{2,3}, Marco Stampanoni^{4,5} & Federica Marone⁴

The rapid diversification of angiosperms through the Early Cretaceous period, between about 130–100 million years ago, initiated fundamental changes in the composition of terrestrial vegetation and is increasingly well understood on the basis of a wealth of palaeobotanical discoveries over the past four decades^{1–5} and their integration with improved knowledge of living angiosperms^{3,6}. Prevailing hypotheses, based on evidence both from living and from fossil plants, emphasize that the earliest angiosperms were plants of small stature^{7–12} with rapid life cycles^{7,8,12,13} that exploited disturbed habitats^{3,9,11,13,14} in open^{3,9,11,13,14}, or perhaps understorey, conditions^{15,16}. However, direct palaeontological data relevant to understanding the seed biology and germination ecology of Early Cretaceous angiosperms are sparse. Here we report the discovery of embryos and their associated nutrient storage tissues in exceptionally well-preserved angiosperm seeds from the Early Cretaceous. Synchrotron radiation X-ray tomographic microscopy of the fossil embryos from many taxa reveals that all were tiny at the time of dispersal. These results support hypotheses based on extant plants that tiny embryos and seed dormancy are basic for angiosperms as a whole^{17,18}. The minute size of the fossil embryos, and the modest nutrient storage tissues dictated by the overall small seed size, is also consistent with the interpretation that many early angiosperms were opportunistic, early successional colonizers of disturbance-prone habitats^{2,15,16}.

As part of a broader survey of Early Cretaceous angiosperm reproductive structures using synchrotron radiation X-ray tomographic microscopy (SRXTM)¹⁹, we analysed the internal structure of mature seeds from about 75 different angiosperm taxa recovered from rich assemblages of angiosperm flowers, fruits and seeds in 11 mesofossil floras from eastern North America and Portugal, ranging in age from Barremian–Aptian to early or middle Albian, about 125–110 million years ago³ (see Methods). SRXTM revealed exquisite preservation of

three-dimensional cellular structure, often including traces of nuclei and subcellular nutritive bodies. In mature fossil fruits and seeds, the seed coat is generally well-developed and cellular preservation is usually excellent. Softer tissues such as embryo and nutrient storage tissues may be degraded or distorted, but of the roughly 250 Early Cretaceous mature seeds examined about half show cellular structure inside the seed coat (Supplementary Table 1). Often only the nutrient storage tissue is preserved, with an empty space at the micropylar end of the seed indicating the maximum size, and former position of the embryo and its immediately surrounding cells. In about 50 seeds, complete or partly preserved embryos occur along with remains of the surrounding nutrient storage tissue. Minimal shrinkage of the seeds during preservation is indicated by the typically straight cell walls and the fact that the nutrient storage tissue often fills out the whole seed volume inside the seed coat.

All Early Cretaceous angiosperm seeds studied here are small (<2.5 mm in maximum dimension²⁰), and in all the fossil seeds in which it can be observed the embryo is tiny. Some embryos have two small cotyledon primordia; in others the cotyledons are not clearly differentiated. None has fully developed cotyledons or a radicle. All were preserved during a dormant phase in their development. Further growth of the embryo inside the seed would be required before germination.

Here we illustrate six different fossils that are representative of the diversity of embryo structure seen among all the specimens studied (Figs 1–3). Three of these fossils can be assigned to extinct genera (*Anacostia*, *Appomattoxia*, *Canrightiopsis*) that have already been described and assessed systematically^{3,21}. The three other taxa (taxon 1, 2 and 3) remain to be described and formally named. Taxon 1 and taxon 3 are isolated exotestal seeds. Taxon 2 is a small, thin-walled seed enclosed in a one-seeded fruiting unit.

In all six kinds of seed, the tiny embryo is surrounded by nutrient storage tissue that occupies the bulk of the space inside the seed coat



Figure 1 | Minute embryos with two cotyledon primordia in Early Cretaceous angiosperms. SRXTM reconstructions of embryos embedded in seeds (a, c, f, h, j) and isolated from seeds (b, d, e, g, i, k). a, b, Exotestal seed and embryo (taxon 1; S170235, Famalicão). c–e, *Canrightiopsis* with seed and embryo (S174005, Famalicão). f, g, *Anacostia* fruit with seed and embryo (PP54021, Kenilworth). h, i, *Appomattoxia* with seed and embryo (PP54064, Puddledock). j, k, Fruit with seed and embryo (taxon 2; PP53991, Kenilworth). Scale bars, 500 μm (a, c, f, h, j), 100 μm (b, d, e, g, i, k).

¹Department of Palaeobiology, Swedish Museum of Natural History, SE-104 05 Stockholm, Sweden. ²Yale School of Forestry and Environmental Studies, 195 Prospect Street, New Haven, Connecticut 06511, USA. ³Department of Earth Science, University of Aarhus, DK-8000 Aarhus, Denmark. ⁴Swiss Light Source, Paul Scherrer Institute, CH-5232 Villigen PSI, Switzerland.

⁵Institute for Biomedical Engineering, ETZ F 85, Swiss Federal Institute of Technology Zurich, Gloriastrasse 35, CH-8092 Zurich, Switzerland.

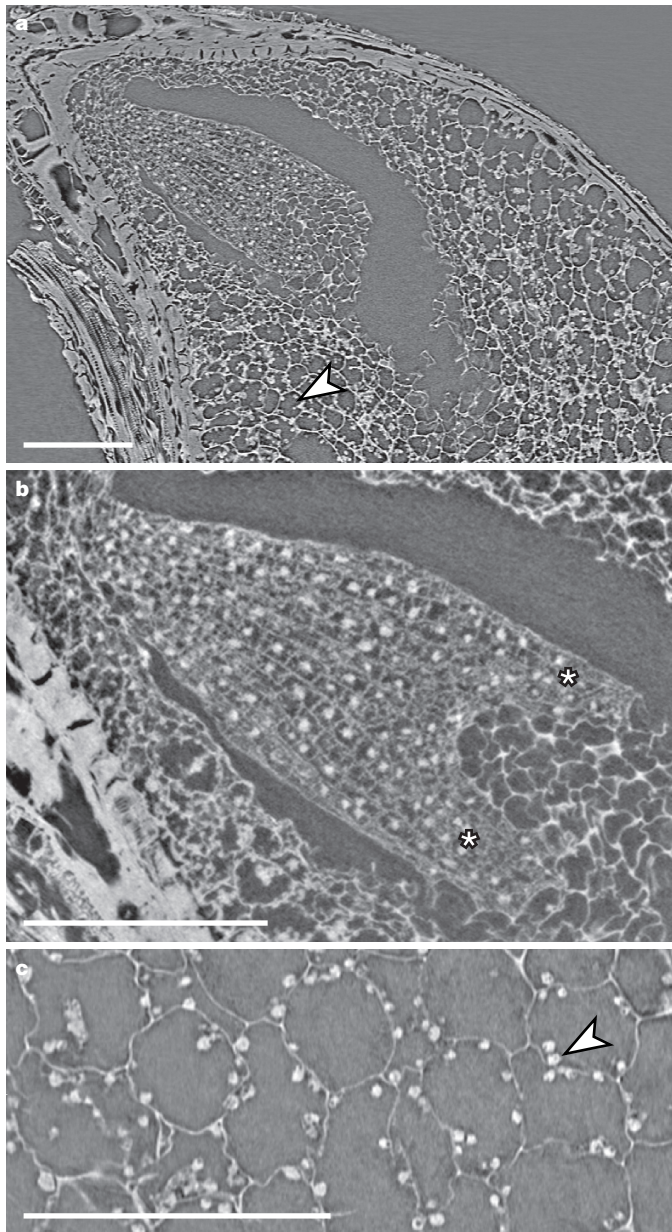


Figure 2 | Cellular preservation of embryos and associated nutrient storage tissue in Early Cretaceous angiosperm seeds. Longitudinal orthoslices through SRXTM volumes. **a**, Apical part of fruit in Fig. 1j (taxon 2) showing embryo and surrounding storage tissue with remains of nutritive bodies (arrow). **b**, Detail of embryo in **a** showing the cotyledon primordia (asterisks) and embryo cells with a central body that may represent remains of the nucleus; thin-walled storage tissue is preserved between the cotyledons. **c**, Details of nutrient storage tissue from an Early Cretaceous exotestal seed (PP53973, Puddledock) with remains of nutritive bodies (arrow). Scale bars, 100 μm .

(Figs 2a and 3), but the size and form of the embryo varies. The cotyledons are not clearly differentiated in taxon 3, and in *Canrightiopsis* and taxon 1 they are rudimentary. In the other three taxa cotyledon primordia are larger. *Canrightiopsis* has the smallest embryo (about 120 μm long) and *Appomattoxia* the largest (about 296 μm long). The embryos of *Anacostia*, taxon 1 and taxon 2 are intermediate in size (*Anacostia* approximately 240 μm long; taxon 1 approximately 250 μm long; taxon 2 about 240 μm long). The embryo in taxon 3 is distinct in being wider than long (about 250 μm wide; 160 μm long). In all seeds examined the embryo size relative to the seed size (two-dimensional area, see Methods) is very small, ranging from 0.015 in taxon 1 to 0.034 in *Anacostia*.

Cellular preservation of the embryos in all six taxa is excellent. Cells are small, rectangular, often elongated parallel to the longitudinal axis and vary in length from 10 to 20 μm . In each cell there is typically a central body about 4–6 μm in diameter (Fig. 2b) that is similar in size and position to the nuclei seen in the embryo cells of extant early diverging angiosperm lineages. The nutrient storage tissue consists of cells that range from about 40 to 70 μm in diameter and have thin, usually straight, walls. Cells in the nutritive storage tissue often contain small rounded structures (Figs 2a, c and 3) that are most probably remains of the protein and lipid bodies that occur in the equivalent seed tissues of many extant angiosperms.

The nutrient storage tissue immediately around the embryo is often partly or fully decomposed, but in seeds with particularly good preservation these cells are usually distinguished by their smaller size, thinner walls and lack of nutritive bodies. Very similar cellular differentiation occurs in the endosperm of modern *Sarcandra* (Fig. 4a, c) and other extant early diverging angiosperm lineages^{22–26}. As in extant taxa, the contents of the cells immediately around the embryo were apparently consumed very early in the development of the young plant.

Taxon 1 (Fig. 1a, b), taxon 3 (Fig. 3) and *Canrightiopsis* (Fig. 1c–e) all have rudimentary or poorly differentiated embryos, as occur in early diverging lineages of living angiosperms (Amborellaceae, Austrobaileyaaceae, Schisandraceae, Nymphaeaceae and Chloranthaceae)^{22–26}, as well as in some eumagnoliids¹⁸. The distinctive exotestal seeds of taxon 1 and taxon 3 are also indicative of a relationship to Schisandraceae or Nymphaeaceae, and the broad embryo of taxon 3 is very similar to the embryos in seeds of extant Nymphaeaceae²⁶.

Canrightiopsis is phylogenetically close to the common ancestor of extant *Ascarina*, *Sarcandra* and *Chloranthus* (Chloranthaceae)²¹. Comparison of the almost spherical *Canrightiopsis* embryo with that of extant *Sarcandra* shows strong similarities and the same cellular features. However, the seeds and embryos of *Canrightiopsis* are much smaller. In *Canrightiopsis* the length of the embryo is about 120 μm (Fig. 1d, e) whereas in the specimen of extant *Sarcandra* illustrated here it is approximately 470 μm (Fig. 4b). Endosperm and perisperm may be difficult to distinguish in mature seeds, but in this case comparison with extant *Sarcandra* strongly suggests that the nutrient storage tissue preserved in *Canrightiopsis* is endosperm.

Anacostia (Fig. 1f, g) and *Appomattoxia* (Fig. 1h, i) are particularly similar in embryo shape and size. Along with taxon 2 (Fig. 1j, k) they have minute embryos with more distinct cotyledons ('underdeveloped linear'²⁷). Embryos of this kind are characteristic of certain lineages among Austrobaileales^{23,24}, eumagnoliids and early diverging eudicots (for example, Ranunculales, Trochodendrales)¹⁸. *Anacostia* and *Appomattoxia* both have abundant monoaperturate pollen on the stigmatic surfaces of their fruits³, making a relationship to eudicots unlikely. Pollen grains of *Anacostia* suggest a relationship to monocots, while other features indicate a position close to Schisandraceae^{3,6}. *Appomattoxia* has features suggesting a relationship to extant Piperales²⁸. In both cases, the minute dicotyledonous embryos are unlike those of the proposed modern relatives, adding further uncertainty to understanding the relationships of these extinct taxa.

Information on the embryos and provisioning of angiosperm seeds from the Early Cretaceous provides new data for assessing their relationships, but also contributes significantly to knowledge of the biology and ecology of early angiosperms. Seed size, based on the new material examined here and previous work, is invariably small^{20,29}, as expected from the small stature documented for some Early Cretaceous angiosperms^{5,9,12} and consistent with the strong relationship between small seed size and small stature seen among living plants³⁰. However, in addition, none of the Early Cretaceous seeds studied here has a fully developed embryo at the time of dispersal. In all cases the embryos are minute and the embryo to seed ratio is much smaller than occurs in

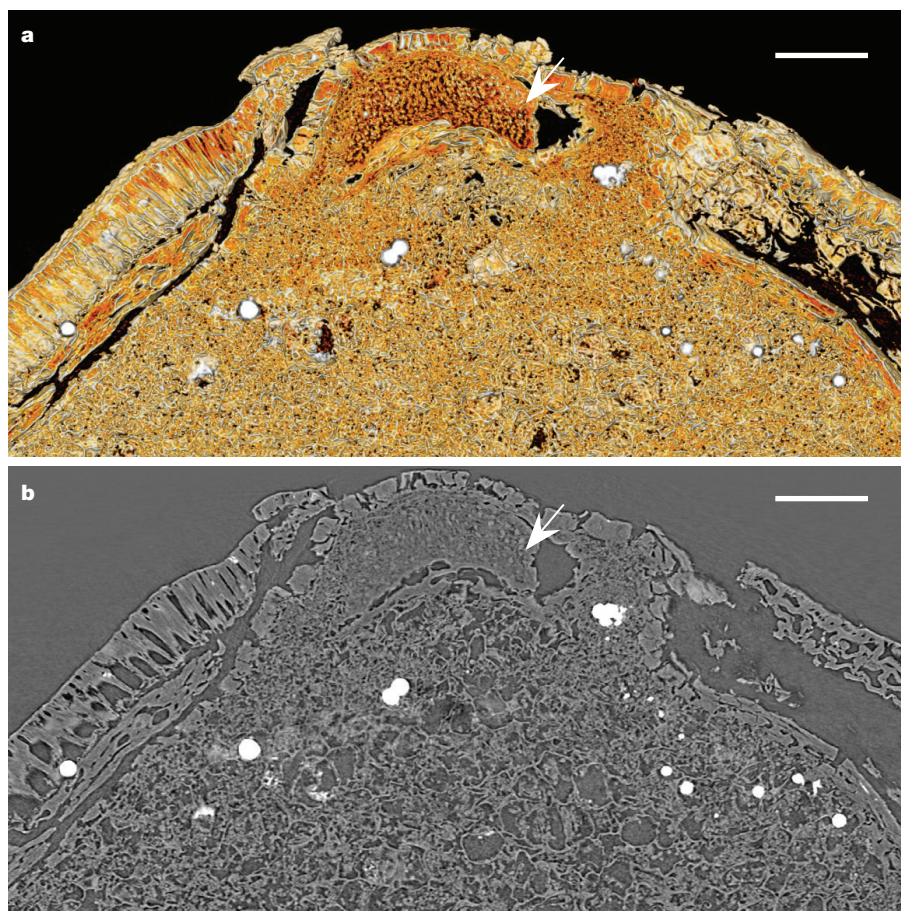


Figure 3 | Minute and broad embryo and associated nutrient storage tissue in an Early Cretaceous seed (taxon 3). Longitudinal two-dimensional SRXTM reconstructions of micropylar region of exotesal seed (S174472, Famalicão 25) showing the broad shape and poorly differentiated embryo (arrow). **a**, Cut volume rendering (between orthoslices 1380–1420) coloured to emphasize the shape and position of embryo. **b**, Single orthoslice (orthoslice 1420) in same position as in **a**. Scale bars, 100 µm.

most extant angiosperms. It is also smaller than the ratio hypothesized for the ancestral angiosperm embryo (0.16 (ref. 17)) by an order of magnitude, emphasizing the additional diversity of extinct taxa close

to the base of the angiosperm phylogenetic tree, and the limitations of inferring ancestral characteristics solely by extrapolation from the features of extant taxa.

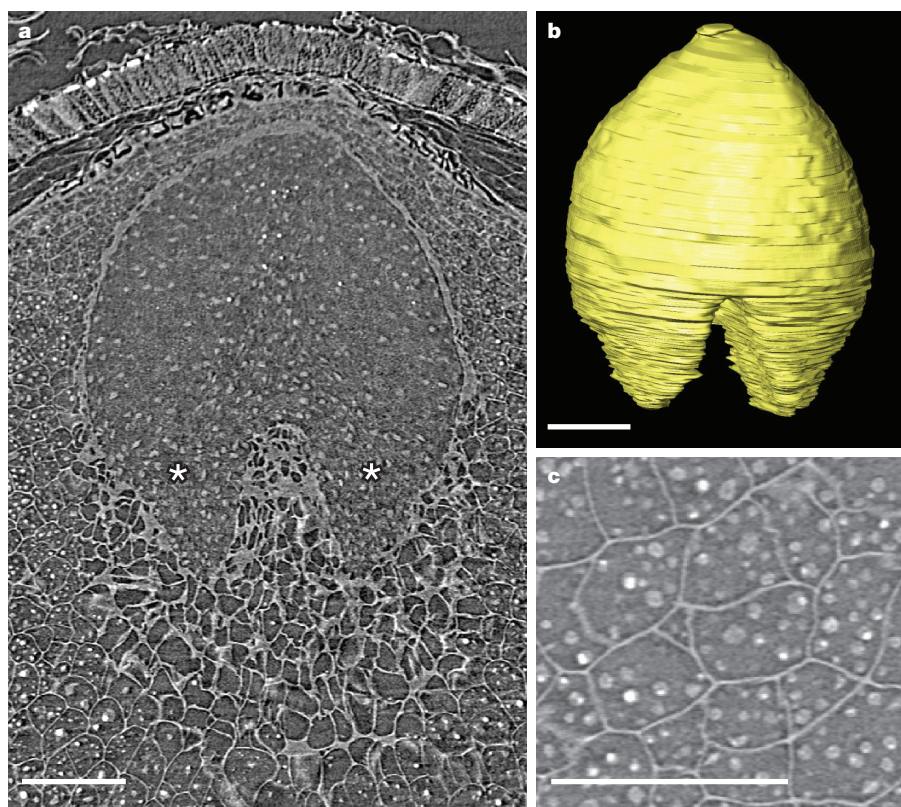


Figure 4 | Embryo and nutrient storage tissue of extant *Sarcandra* (Chloranthaceae). Two- (**a**, **c**) and three-dimensional (**b**) SRXTM reconstructions. **a**, Longitudinal orthoslice through seed showing rudimentary embryo with two cotyledon primordia (asterisks) embedded in copious nutrient storage tissue (endosperm); cells in the vicinity of the embryo lack the nutritive bodies that are abundant in other endosperm cells. **b**, Surface rendering of embryo showing the two small cotyledon primordia. **c**, Detail of endosperm with nutritive bodies (protein and lipids). Scale bars, 100 µm.

Seed dormancy associated with the minute fossil embryos ensured that the seeds of early angiosperms could survive until conditions for germination and seedling establishment were favourable. However, the tiny embryo size and modest nutrient reserves were also an intrinsic developmental constraint on the rapidity with which early angiosperms could germinate in response to short-lived moisture availability. Early angiosperms would have been unable to match the very rapid germination of many angiosperms that evolved later and ultimately proved even more effective in exploiting ephemeral ecological opportunities.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 September; accepted 5 November 2015.

Published online 16 December 2015.

- Hughes, N. F. *The Enigma of Angiosperm Origins* (Cambridge Univ. Press, 1994).
- Doyle, J. A. & Hickey, L. J. in *Origin and Early Evolution of Angiosperms* (ed. Beck, C. B.) 139–206 (Columbia Univ. Press, 1976).
- Friis, E. M., Crane, P. R. & Pedersen, K. R. *Early Flowers and Angiosperm Evolution* (Cambridge Univ. Press, 2011).
- Dilcher, D. L. Early angiosperm reproduction: an introductory report. *Rev. Palaeobot. Palynol.* **27**, 291–328 (1979).
- Sun, G. et al. Archaeofractaceae, a new basal angiosperm family. *Science* **296**, 899–904 (2002).
- Doyle, J. A. & Endress, P. K. Integrating Early Cretaceous fossils into the phylogeny of living angiosperms: Magnoliidae and eudicots. *J. Syst. Evol.* **48**, 1–35 (2010).
- Stebbins, G. L. The probable growth habit of the earliest flowering plants. *Ann. Mo. Bot. Gard.* **52**, 457–468 (1965).
- Stebbins, G. L. in *Origin and Early Evolution of Angiosperms* (ed. Beck, C. B.) 300–311 (Columbia Univ. Press, 1976).
- Taylor, D. W. & Hickey, L. J. in *Flowering Plant Origin, Evolution and Phylogeny* (eds Taylor, D. W. & Hickey, L. J.) 232–266 (Chapman & Hall, 1996).
- Wing, S. L. & Boucher, L. D. Ecological aspects of the Cretaceous flowering plant radiation. *Annu. Rev. Earth Planet. Sci.* **26**, 379–421 (1998).
- Friis, E. M., Pedersen, K. R. & Crane, P. R. Diversity in obscurity: fossil flowers and the early history of angiosperms. *Phil. Trans. R. Soc. B* **365**, 369–382 (2010).
- Jud, N. A. Fossil evidence for a herbaceous diversification of early eudicot angiosperms during the Early Cretaceous. *Proc. R. Soc. B* **282**, 20151045 (2015).
- Royer, D. L., Miller, I. M., Peppe, D. J. & Hickey, L. J. Leaf economic traits from fossils support a weedy habit for early angiosperms. *Am. J. Bot.* **97**, 438–445 (2010).
- Taylor, D. W. & Hickey, L. J. An aptian plant with attached leaves and flowers: implications for angiosperm origin. *Science* **247**, 702–704 (1990).
- Feild, T. S., Arens, A. C., Doyle, J. A., Dawson, T. E. & Donoghue, M. J. Dark and disturbed: a new image of early angiosperm ecology. *Paleobiology* **30**, 82–107 (2004).
- Lee, A. P., Upchurch, G., Jr, Murchie, E. H. & Lomax, B. H. Leaf energy balance modelling as a tool to infer habitat preference in the early angiosperms. *Proc. R. Soc. B* **282**, 20143052 (2015).
- Forbis, T. A., Floyd, S. K. & de Queiroz, A. The evolution of embryo size in angiosperms and other seed plants: implications for the evolution of seed dormancy. *Evolution* **56**, 2112–2125 (2002).
- Baskin, C. C. & Baskin, J. M. *Seeds, Ecology, Biogeography, and Evolution of Dormancy and Germination* 2nd edn, 1–1586 (Academic, 2014).
- Friis, E. M., Marone, F., Pedersen, K. R., Crane, P. R. & Stampanoni, M. Three-dimensional visualization of fossil flowers, fruits, seeds and other plant remains using synchrotron radiation X-ray tomographic microscopy (SRXTM): new insights into Cretaceous plant diversity. *J. Paleontol.* **88**, 684–701 (2014).
- Eriksson, O., Friis, E. M., Pedersen, K. R. & Crane, P. R. Seed size and dispersal systems of Early Cretaceous angiosperms from Famalicão, Portugal. *Int. J. Plant Sci.* **161**, 319–329 (2000).
- Friis, E. M., Grimm, G. W., Mendes, M. M. & Pedersen, K. R. *Canrightiopsis*, a new Early Cretaceous fossil with *Clavatipollenites*-type pollen bridge the gap between extinct *Canrightia* and extant Chloranthaceae. *Grana* **54**, 184–212 (2015).
- Floyd, S. K. & Friedman, W. E. Evolution of endosperm developmental patterns among basal flowering plants. *Int. J. Plant Sci.* **161**, S57–S81 (2000).
- Friedman, W. E. & Bachelier, J. B. Seed development in *Trimenia* (Trimeniaceae) and its bearing on the evolution of embryo-nourishing strategies in early flowering plant lineages. *Am. J. Bot.* **100**, 906–915 (2013).
- Floyd, S. K. & Friedman, W. E. Developmental evolution of endosperm in basal angiosperms: Evidence from *Amborella* (Amborellaceae), *Nuphar* (Nymphaeaceae), and *Illicium* (Illiciaceae). *Plant Syst. Evol.* **228**, 153–169 (2001).
- Tobe, H., Jaffre, T. & Raven, P. H. Embryology of *Amborella* (Amborellaceae): descriptions and polarity of character states. *J. Plant Res.* **113**, 271–280 (2000).
- Povilus, R. A., Losada, J. M. & Friedman, W. E. Floral biology and ovule and seed ontogeny of *Nymphaea thermarum*, a water lily at the brink of extinction with potential as a model system for basal angiosperms. *Ann. Bot. (Lond.)* **115**, 211–226 (2015).
- Baskin, C. C. & Baskin, J. M. A revision of Martin's seed classification system, with particular reference to his dwarf-seed type. *Seed Sci. Res.* **17**, 11–20 (2007).
- Friis, E. M., Pedersen, K. R. & Crane, P. R. *Appomattoxia ancistrophora* gen. et sp. nov., a new Early Cretaceous plant with similarities to *Circaeaster* and extant Magnoliidae. *Am. J. Bot.* **82**, 933–943 (1995).
- Eriksson, O., Friis, E. M. & Löfgren, P. Seed size, fruit size, and dispersal systems in angiosperms from the Early Cretaceous to the Late Tertiary. *Am. Nat.* **156**, 47–58 (2000).
- Moles, A. T. et al. A brief history of seed size. *Science* **307**, 576–580 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Lindström for assistance with the SRXTM analyses. Research reported here was supported by the Swedish Research Council, the Edward P. Bass Distinguished Visiting Fellowship and by the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement n. 312284 (for CALIPSO) for the SRXTM analyses at the SLS.

Author Contributions E.M.F., K.R.P. and P.R.C. collected and prepared the fossil material for analyses. The measurements and reconstructions were performed by E.M.F. F.M. and M.S. developed the algorithms for the analyses and enhanced the measurements. The paper was jointly prepared by the authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.M.F. (else.marie.friis@nrm.se).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

The fossil seeds studied here were isolated from 11 mesofossil floras preserved in soft unconsolidated sediments from eastern North America (Kenilworth, Maryland; Dutch Gap and Puddledock, Virginia) and Portugal (Arazede, Buarcos, Catefica, Famalicão, Juncal-Chicalhão, Torres Vedras, Vale de Água, Vila Verde) that range from Barremian–Aptian to early or middle Albian in age^{3,21,28,31}. Mesofossils preserved in these floras are often exquisitely preserved in three dimensions as charcoallified or lignitic specimens and include complete and fragmentary flowers, as well as abundant fruits and seeds. Fossils were isolated from the sediments by sieving in water, remaining mineral matrix was removed using hydrofluoric and hydrochloric acids, and the fossils were then rinsed in water and air-dried. Many specimens of mature seeds, from the full range of taxa preserved, were analysed using SRXTM. Six fossils representative of the material examined were selected to illustrate common features of embryos and nutritive storage tissues. Specimens examined with SRXTM were mounted on brass stubs with nail polish and analysed at the TOMCAT beamline³² at the Swiss Light Source, Villigen, Switzerland. For optimized contrast, measurements were made at 10 keV. For each data set, 1,501 projections equiangularly spaced over 180° were acquired. The transmitted and refracted X-ray radiation was converted to visible light by a thin scintillating screen (20 µm thick LAG:Ce or 5.9 µm thick LSO:Tb depending on the spatial resolution required), magnified by ×10 and ×20 objective lenses for overviews, and ×40 objective lens for details, and digitized by a charge coupled device (PCO.2000) or a scientific complementary metal-oxide–semiconductor (sCMOS) (PCO.edge) camera. The sample-detector distance was of the order of few millimetres. The raw projections were dark- and flat-field corrected and subsequently reconstructed using an efficient algorithm based on the Fourier method with regridding³³. The resulting volumetric data have voxel sizes of 0.65–0.74, 0.325

and 0.1625 µm for measurements done with the ×10, ×20 and ×40 objectives respectively.

To boost contrast in the detailed scan of specimen PP53991 (Fig. 2b, c), before tomographic reconstruction, the corrected projections were phase retrieved according to the single distance algorithm in ref. 34.

Embryo tissue was identified in the reconstructed SRXTM orthoslices and Avizo software was used to manually label individual slices to generate the three-dimensional embryo shapes. To illustrate the relationship of seed and embryo volume, the embryo surface was coloured yellow and the three-dimensional shape of the seeds/fruits shown by transparent voltex rendering in green (Fig. 1). The two-dimensional area of embryo and seed inside the integuments was measured in pixels on longitudinal sections through the middle of the seeds and embryos using the free software Fiji³⁵, resulting in an embryo to seed ratio comparable to that published by others¹⁷.

A list of the mature seeds analysed here is available in Supplementary Table 1. The fossil material is stored in the palaeobotanical collections of the Swedish Museum of Natural History, Stockholm (S), and the Field Museum, Chicago (PP). Raw data from the SRXTM are stored at the Swedish Museum of Natural History.

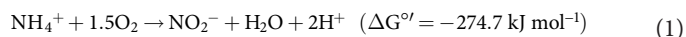
31. Friis, E. M., Crane, P. R. & Pedersen, K. R. *Anacostia*, a new basal angiosperm from the Early Cretaceous of North America and Portugal with monocolpate/trichotomocolpate pollen. *Grana* **36**, 225–244 (1997).
32. Stampanoni, M. *et al.* in *Developments in X-Ray Tomography V* Vol. 6318 (ed Bonse, U.) (International Society for Optical Engineering, 2006).
33. Marone, F. & Stampanoni, M. Regridding reconstruction algorithm for real-time tomographic imaging. *J. Synchrotron Radiat.* **19**, 1029–1037 (2012).
34. Paganin, D., Mayo, S. C., Gureyev, T. E., Miller, P. R. & Wilkins, S. W. Simultaneous phase and amplitude extraction from a single defocused image of a homogeneous object. *J. Microsc.* **206**, 33–40 (2002).
35. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**, 676–682 (2012).

Complete nitrification by a single microorganism

Maartje A. H. J. van Kessel¹, Daan R. Speth¹, Mads Albertsen², Per H. Nielsen², Huub J. M. Op den Camp¹, Boran Kartal^{1,3}, Mike S. M. Jetten^{1,4} & Sebastian Lüscher¹

Nitrification is a two-step process where ammonia is first oxidized to nitrite by ammonia-oxidizing bacteria and/or archaea, and subsequently to nitrate by nitrite-oxidizing bacteria. Already described by Winogradsky in 1890¹, this division of labour between the two functional groups is a generally accepted characteristic of the biogeochemical nitrogen cycle². Complete oxidation of ammonia to nitrate in one organism (complete ammonia oxidation; comammox) is energetically feasible, and it was postulated that this process could occur under conditions selecting for species with lower growth rates but higher growth yields than canonical ammonia-oxidizing microorganisms³. Still, organisms catalysing this process have not yet been discovered. Here we report the enrichment and initial characterization of two *Nitrospira* species that encode all the enzymes necessary for ammonia oxidation via nitrite to nitrate in their genomes, and indeed completely oxidize ammonium to nitrate to conserve energy. Their ammonia monooxygenase (AMO) enzymes are phylogenetically distinct from currently identified AMOs, rendering recent acquisition by horizontal gene transfer from known ammonia-oxidizing microorganisms unlikely. We also found highly similar *amoA* sequences (encoding the AMO subunit A) in public sequence databases, which were apparently misclassified as methane monooxygenases. This recognition of a novel *amoA* sequence group will lead to an improved understanding of the environmental abundance and distribution of ammonia-oxidizing microorganisms. Furthermore, the discovery of the long-sought-after comammox process will change our perception of the nitrogen cycle.

Nitrification, the aerobic oxidation of ammonium to nitrate is divided into two subsequent reactions: ammonium oxidation to nitrite (equation (1)) and nitrite oxidation to nitrate (equation (2)). These two reactions are catalysed by physiologically distinct clades of microorganisms.



Even though the existence of a single microorganism capable of oxidizing ammonium to nitrate (equation (3)) was not previously reported, it was proposed that such a microorganism could have a competitive advantage in biofilms and other microbial aggregates with low substrate concentrations³.

In this study, to characterize the microorganisms responsible for nitrogen transformations in an ammonium-oxidizing biofilm, we sampled the anaerobic compartment of a trickling filter connected to a recirculation aquaculture system⁴ with an ammonium effluent of less than 100 μM . To enrich for the N-cycling community, a bioreactor was inoculated and supplied with low concentrations of ammonium, nitrite and nitrate under hypoxic conditions ($\leq 3.1 \mu\text{M O}_2$). Within 12 months, we obtained a stable enrichment culture that efficiently

removed ammonium and nitrite from the medium (Extended Data Fig. 1). The culture showed anaerobic ammonium-oxidizing (anammox) activity (Fig. 1a), and fluorescence *in situ* hybridization (FISH) revealed that anammox organisms of the genus *Brocadia* constituted approximately 45% of all FISH-detectable bacteria. Surprisingly, *Nitrospira*-like nitrite-oxidizing bacteria accounted for approximately 15% of the community and co-occurred with the *Brocadia* species in flocs (Fig. 2a). This tight clustering with anammox bacteria was unexpected as both microorganisms require nitrite for growth. Together with the presence of *Nitrospira* at very low oxygen concentrations, this indicated that there could be a functional link between these organisms.

To determine the function of *Nitrospira* in the community, we extracted and sequenced total DNA from the enrichment culture biomass. In total 4.95 gigabase pairs of trimmed metagenomic sequence were obtained and used for *de novo* assembly. By differential coverage and sequence composition-based binning⁵ it was possible to extract high-quality draft genomes of two *Nitrospira* species. The two strains had genomic pairwise average nucleotide identities (ANI)⁶ of 75% and thus clearly represented different species (*Nitrospira* sp.1 and sp.2, Extended Data Fig. 2 and Extended Data Table 1). Surprisingly, both genomes contained the full set of AMO and hydroxylamine dehydrogenase (HAO) genes for ammonia oxidation, in addition to the nitrite oxidoreductase (NXR) subunits necessary for nitrite oxidation in *Nitrospira*⁷. In both species all these genes were localized on a single contiguous genomic fragment, along with general housekeeping genes that allowed reliable phylogenetic assignment. Consequently, these *Nitrospira* species had the genetic potential for the complete oxidation of ammonia to nitrate. No AMO of canonical ammonia-oxidizing bacteria or archaea could be detected in the trimmed metagenomic reads or by *amoA*-specific PCR^{8,9} on DNA extracted from reactor biomass, and no other indications for the presence of ammonia-oxidizing microorganisms were found in the metagenome or by FISH analyses. The AMO structural genes (*amoCAB*) of both *Nitrospira* species, along with the putative additional AMO subunits *amoEDD2*^{10,11}, formed one gene cluster with *haoAB-cycAB* (encoding HAO, the putative membrane anchor protein HaoB, electron transfer protein cytochrome *c*₅₅₄ and quinone-reducing cytochrome *c*_{m552}, respectively)¹² and showed highest similarities to their counterparts in betaproteobacterial ammonia-oxidizing bacteria (60% average amino acid identity to the *Nitrosomonas europaea* genes; Fig. 3 and Supplementary Table 1). The same genomic region also contained genes for copper and haem transport, cytochrome *c* biosynthesis, and iron storage. These accessory genes were highly conserved in ammonia-oxidizing bacteria but not in other *Nitrospira*^{7,13}, indicating their involvement in AMO and HAO biosynthesis or activation. *Nitrospira* sp.1 encoded three discrete *amoC* genes, one of which was clustered with a second, almost identical copy of *amoA* (97.7% amino acid identity). *Nitrospira* sp.2 lacked the second *amoA*, but contained four additional *amoC* and a second *haoA* gene (Supplementary Table 1). Unlike other *Nitrospira*^{7,13}, both species lacked enzymes for assimilatory nitrite reduction, indicating adaptation to ammonium-containing

¹Department of Microbiology, IWW, Radboud University, Heyendaalseweg 135, 6525 AJ Nijmegen, the Netherlands. ²Center for Microbial Communities, Department of Chemistry and Bioscience, Aalborg University, Fredrik Bajers Vej 7H, 9220 Aalborg, Denmark. ³Laboratory for Microbiology, University of Gent, K. L. Ledeganckstraat 35, 9000 Gent, Belgium. ⁴Department of Biotechnology, TU Delft, Julianalaan 67, 2628 BC Delft, the Netherlands.

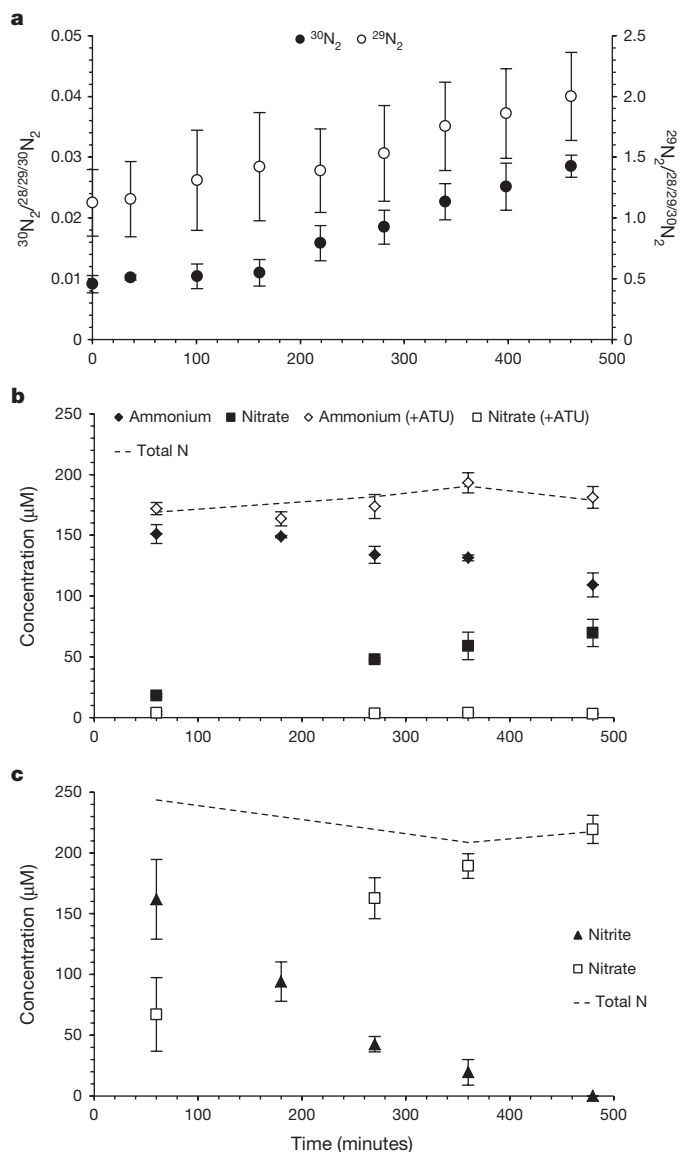


Figure 1 | Ammonium oxidation by the enrichment culture. **a**, $^{29}\text{N}_2$ (open circles) and $^{30}\text{N}_2$ (filled circles) production from $^{15}\text{NH}_4^+$ by the enrichment culture. **b**, Ammonium (diamonds) oxidation to nitrate (squares) in aerobic batch incubations in the absence (filled symbols) and presence (open symbols) of ATU. Nitrite concentrations were below the detection limit ($<5\mu\text{M}$) at all time points. **c**, Nitrite (triangles) oxidation to nitrate (squares) in aerobic batch incubations. In **b** and **c**, total nitrogen balances are indicated (dashed lines). Symbols in all plots represent averages of three individual experiments. Ammonium concentrations were determined in single measurements, other compounds in triplicate. Error bars represent standard deviations of three biological replicates.

habitats. For ammonium uptake, they encoded low-affinity Rh-type transporters most closely related to Rh50 found in *Nitrosomonas europaea*¹⁴, in contrast to most ammonia-oxidizing and nitrite-oxidizing bacteria that have the high-affinity AmtB-type proteins. Both species encoded ureases and the corresponding ABC transport systems, indicating that urea could be used as an alternative ammonium source. Interestingly, *Candidatus Nitrospira inopinata*, the moderately thermophilic ammonia-oxidizing *Nitrospira* described by Daims *et al.*¹⁵, encoded a similar set of AMO, HAO and urease proteins, and also lacked genes for assimilatory nitrite reduction. Unlike the two species described here, however, it contained a periplasmic cytochrome *c* nitrite reductase (NrfA) that could allow it to conserve energy by dissimilatory nitrite reduction to ammonium (DNRA), but might also provide ammonium for assimilation. The evolutionary divergence of

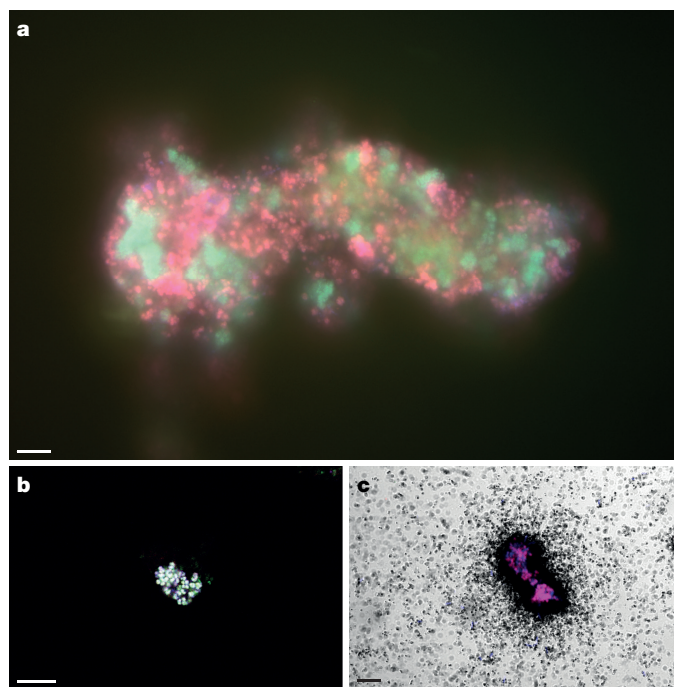


Figure 2 | In situ detection of *Nitrospira* and their ammonia-oxidizing capacity. **a**, Co-aggregation of *Nitrospira* and *Brocadia* in the enrichment. Cells are stained by FISH with probes for all bacteria (EUB338mix, blue), and specific for *Nitrospira* (Ntspa712, green, resulting in cyan) and anammox bacteria (Amx820, red, resulting in magenta). **b**, AMO labelling by FTCP (green). *Nitrospira* was counterstained by FISH (probes Ntspa662 (blue) and Ntspa476 (red), resulting in white). **c**, Ammonium-dependent CO_2 fixation by *Nitrospira* shown by FISH-MAR. Silver grain deposition (black) above cell clusters indicates $^{14}\text{CO}_2$ incorporation. *Nitrospira* was stained by FISH (probes Ntspa476 (red) and Ntspa662 (blue), resulting in magenta). Images in **b** and **c** are representative of two individual experiments, with three (**b**) or two (**c**) technical replicates each. Scale bars in all panels represent $10\mu\text{m}$.

these organisms was also reflected in the low ANI values of 70.3–71.6% between *Candidatus N. inopinata* and the two species described here. Concerning their genetic repertoire for nitrite oxidation, sp.2 had four almost identical ($>99\%$ amino acid identity) NXR alpha and beta (NxrAB) subunits. Sp.1 had two *nxrAB* copies encoding identical NxrB subunits, but NxrA subunits with amino acid identities of 89.6%, which were separated into distinct clusters in phylogenetic analyses. One homologue branched with sequences from *Nitrospira moscoviensis*, while the other formed a novel sequence cluster together with the sequences from sp.2 (Extended Data Fig. 3).

To ascertain that ammonia oxidation occurred under hypoxic conditions in the enrichment culture, we supplied the bioreactor with ^{15}N -labelled ammonium. While the anammox bacteria consumed $^{15}\text{NH}_4^+$ and converted it into $^{29}\text{N}_2$, a steady increase of $^{30}\text{N}_2$ was also observed (Fig. 1a). This formation of $^{30}\text{N}_2$ could only be explained by the production of ^{15}N -labelled nitrite derived through aerobic ammonium oxidation. As metagenomic analyses confirmed that the *Nitrospira* species were the only organism in the enrichment harbouring AMO and HAO, this clearly showed that they were able to perform this reaction even at O_2 concentrations lower than $3.1\mu\text{M}$. To unambiguously link this activity to *Nitrospira*, we visualized the AMO protein *in situ* using batch incubations with reactor biomass and FTCP (fluorescein thiocarbamoylpropargylamine), a fluorescently labelled acetylene analogue that acts as suicide substrate for AMO¹⁶ and covalently binds to the enzyme¹⁷. When counterstained with *Nitrospira*-specific FISH probes, including a newly designed probe specifically targeting the 16S ribosomal RNA-defined phylogenetic group comprising spp.1 and 2 (Extended Data Table 2 and Extended Data Fig. 4), strong FTCP labelling of *Nitrospira* cells was observed, providing strong support for

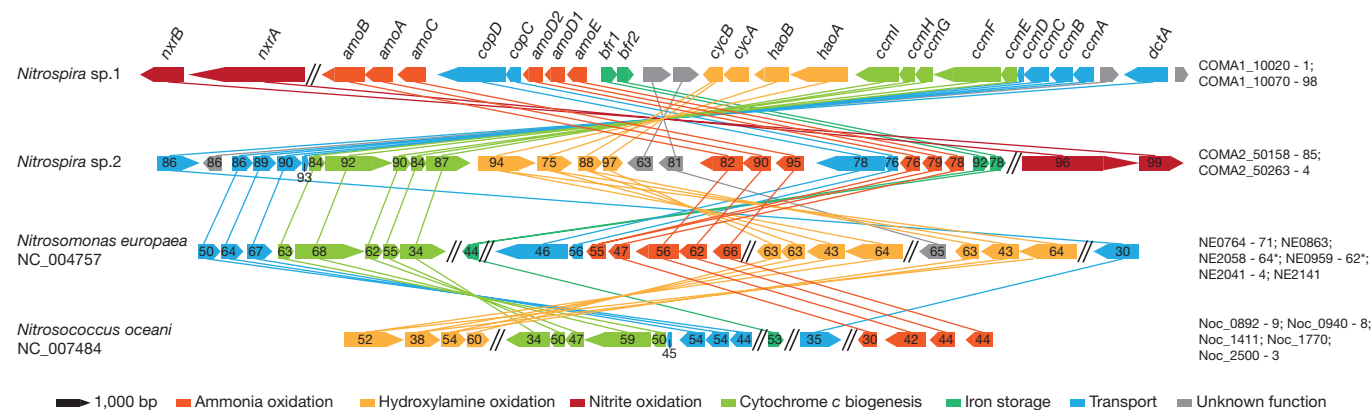


Figure 3 | Schematic illustration of the AMO genomic region in *Nitrospira* and selected ammonia-oxidizing bacteria. The AMO locus in *Nitrospira* sp.1 in comparison to sp.2 and the beta- and gammaproteobacterial ammonia-oxidizing bacteria *Nitrosomonas europaea* and *Nitrosococcus oceanii*, respectively. The position of NXR on the AMO-containing *Nitrospira* contigs is also indicated. Homologous genes are connected by lines. Functions of the encoded proteins are

represented by colour, the arrow shows direction of transcription. Numbers specify amino acid identities to *Nitrospira* sp.1. Parallel double lines designate a break in locus organization. Locus tags for each organism are listed on the right. Genes are drawn to scale. *amo*, ammonia monooxygenase; *bfr*, bacterioferritin; *ccm*, cytochrome c biogenesis; *cop*, copper transport; *cyc*, cytochrome c; *dct*, sodium:dicarboxylate symporter; *hao*, hydroxylamine dehydrogenase; *nrx*, nitrite oxidoreductase.

the presence of the ammonia-oxidizing enzyme at the single-cell level (Fig. 2b and Extended Data Fig. 5).

Batch incubations were performed at ambient oxygen concentrations to determine conversion rates of ammonium and nitrite, the level of inhibition by allylthiourea (ATU; a potent inhibitor of bacterial ammonia oxidation^{18,19}), and the use of urea as ammonium source for nitrification. Flocs were mechanically disrupted to ensure complete exposure of the biomass to oxygen, which inhibits the anammox and denitrification processes^{20,21}. This inhibition was confirmed by the lack of labelled N_2 formation in incubations with $^{15}NH_4^+$. In these incubations (Fig. 1 and Extended Data Fig. 6), the culture oxidized ammonium ($6.0 \pm 1.0 \mu M h^{-1} NH_4^+$) and nitrite ($23 \pm 4.7 \mu M h^{-1} NO_2^-$) to nitrate. ATU selectively inhibited ammonia oxidation, but did not affect nitrite oxidation rates. Urea was converted to ammonium, which was subsequently oxidized to nitrate ($7.8 \pm 1.1 \mu M h^{-1} NO_3^-$), suggesting that these *Nitrospira* species were capable of using urea as source of ammonia to drive nitrification, as was also reported for some ammonia-oxidizing archaea²² and bacteria²³. This trait could enable them to thrive in environments like fertilized soils, wastewater treatment plants, and many aquatic systems where urea is often present at micromolar levels²⁴. However, it should be noted that the two *Nitrospira* spp. were not the only organisms in the enrichment culture that encoded ureases.

To investigate substrate-dependent inorganic carbon fixation as a proxy for energy conservation from ammonia and nitrite oxidation, we used FISH in combination with microautoradiography (FISH-MAR)²⁵. Aerobic incubations with mechanically disrupted flocs were performed in the presence of 500 μM ammonium, 500 μM ammonium with 100 μM ATU, or 500 μM nitrite. *Nitrospira* incorporated carbon from ^{14}C -labelled bicarbonate in the presence of either ammonium or nitrite, and ammonia-dependent carbon fixation was strongly inhibited by the addition of ATU (Fig. 2c and Extended Data Fig. 7). Only flocs containing *Nitrospira* were labelled during all incubations, suggesting that these were the only chemolithoautotrophic nitrifying organisms present in the culture and indeed could conserve energy from the oxidation of ammonia and nitrite.

In 16S rRNA-based phylogenetic analyses, the two ammonia-oxidizing *Nitrospira* species from our enrichment culture formed two separate lineages within one strongly supported sequence cluster affiliated with *Nitrospira* sublineage II²⁶ (Extended Data Fig. 4). They both grouped with highly similar sequences (>99% nucleotide identity) from a diverse range of habitats, including soil, groundwater, recirculation aquaculture systems, wastewater treatment plants and drinking water distribution systems. The formation of distinct clusters containing sp.1 and sp.2 indicated that the last common ancestor encoded genes for

complete nitrification and that this pathway might be conserved in most organisms affiliated with this sequence group.

To explore the environmental relevance of these *Nitrospira*, we searched the NCBI nr database²⁷ for closely related *amoA* genes. Surprisingly, we found the *AmoA* proteins of the two *Nitrospira* species to be phylogenetically divergent from the described bacterial *AmoA* sequences. *Nitrospira* sp.2 *AmoA* was 97–98% identical to the so-called ‘unusual’ methane monooxygenase (PMO) proteins of *Crenothrix polyspora*²⁸. The two *AmoA* copies from *Nitrospira* sp.1 had lower similarities to *Crenothrix* *PmoA* (90–91% identity), but also affiliated with this group (Fig. 4). Sequences within this group cannot be amplified by standard *amoA* primers, but only by *pmoA* primers when used at reduced stringency²⁹. Therefore the public databases only contain few closely related sequences, which were mainly derived from habitats studied for their bacterial methane-oxidizing communities. Highly similar sequences derived from wastewater treatment plants and drinking water systems, however, indicated occurrence of ammonia-oxidizing *Nitrospira* in a range of engineered and natural environments. We furthermore screened all publicly available shotgun data sets on MG-RAST³⁰. Indeed, 168 metagenomes (out of 6,255) and 28 metatranscriptomes (out of 1,051) contained at least two reads affiliated with this *amoA* group, yielding a total of 3,727 reads that were obtained mainly from soil, sediments and wastewater treatment plants (Extended Data Table 3). Thus, our results showed that the *Crenothrix* sequence group consists of so far unrecognized AMO sequences overlooked in nitrification studies based on *amoA* gene detection. Based on these findings, it is highly likely that the PCR-based determination of the *Crenothrix pmoA* gene from an environmental sample²⁸ was erroneous, and this cluster only contains genes encoding AMOs. Nevertheless, with the currently available information it cannot be excluded that certain *Crenothrix* species attained an *amoA* gene through lateral gene transfer and use the encoded protein as a surrogate PMO.

In conclusion, here we demonstrated the existence of complete nitrification in a single organism (comammox) and identified two *Nitrospira* species capable of catalysing this process (equation (3)). In 16S rRNA or *amoA/pmoA*-based studies these organisms would have been classified as nitrite-oxidizing or methane-oxidizing bacteria, respectively. Hence, our results show that a whole group of ammonia-oxidizing organisms was previously overlooked. Our findings furthermore disprove the long-held assumption that nitrification (ammonia oxidation via nitrite to nitrate) is catalysed by two distinct functional groups, thus redefining a key process of the biogeochemical nitrogen cycle.

Based on their physiology, differences in genome content, and separation in different phylogenetic groups in 16S rRNA-based analyses,

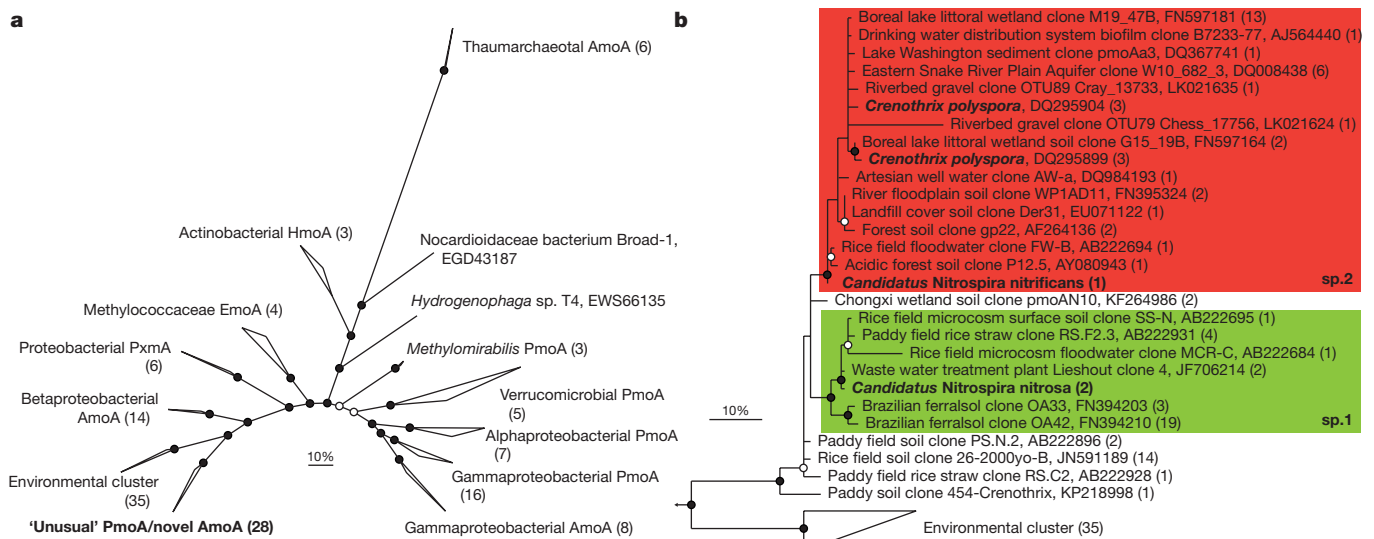


Figure 4 | Phylogenetic analysis of the AmoA/PmoA sequence family. Bayesian inference tree (s.d. = 0.01) showing the affiliation of the *Nitrospira* AmoA. Posterior probabilities $\geq 70\%$ and $\geq 90\%$ are indicated by open and filled circles, respectively. Scale bars indicate 10% sequence divergence. **a**, Radial tree indicating the localization of the novel AmoA/'unusual' PmoA sequence group in relation to the main functional groups within the sequence family. Numbers in brackets indicate sequences per group (137 sequences in total). Amo, ammonia

we propose tentative names for both *Nitrospira* species present in our enrichment: *Candidatus Nitrospira nitrosa* (etymology: L. fem. adj. *nitrosa*, nitrous; the nitrite and nitrate forming *Nitrospira*) for sp.1 and *Candidatus Nitrospira nitrificans* (N.L. part. adj. *nitrificans*, nitrifying; the nitrifying *Nitrospira*) for sp.2. Both species are chemolithoautotrophic and fully oxidize ammonia via nitrite to nitrate.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 August; accepted 18 November 2015.

Published online 26 November 2015.

- Winogradsky, S. Recherches sur les organismes de la nitrification. *Ann. Inst. Pasteur (Paris)* **4**, 213–231 (1890).
- Vlaeminck, S. E., Hay, A. G., Maignien, L. & Verstraete, W. In quest of the nitrogen oxidizing prokaryotes of the early Earth. *Environ. Microbiol.* **13**, 283–295 (2011).
- Costa, E., Pérez, J. & Kreft, J. U. Why is metabolic labour divided in nitrification? *Trends Microbiol.* **14**, 213–219 (2006).
- Crab, R., Avnimelech, Y., Defoirdt, T., Bossier, P. & Verstraete, W. Nitrogen removal techniques in aquaculture for a sustainable production. *Aquaculture* **270**, 1–14 (2007).
- Albertsen, M. *et al.* Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature Biotechnol.* **31**, 533–538 (2013).
- Richter, M. & Rosselló-Móra, R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl Acad. Sci. USA* **106**, 19126–19131 (2009).
- Lücker, S. *et al.* A *Nitrospira* metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proc. Natl Acad. Sci. USA* **107**, 13479–13484 (2010).
- Rothauwe, J. H., Witzel, K. P. & Liesack, W. The ammonia monooxygenase structural gene *amoA* as a functional marker: molecular fine-scale analysis of natural ammonia-oxidizing populations. *Appl. Environ. Microbiol.* **63**, 4704–4712 (1997).
- Könneke, M. *et al.* Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**, 543–546 (2005).
- El Sheikh, A. F., Poret-Peterson, A. T. & Klotz, M. G. Characterization of two new genes, *amoR* and *amoD*, in the *amo* operon of the marine ammonia oxidizer *Nitrosococcus oceanus* ATCC 19707. *Appl. Environ. Microbiol.* **74**, 312–318 (2008).
- Berube, P. M. & Stahl, D. A. The divergent *AmoC*₃ subunit of ammonia monooxygenase functions as part of a stress response system in *Nitrosomonas europaea*. *J. Bacteriol.* **194**, 3448–3456 (2012).

monooxygenase; Emo, ethane monooxygenase; Hmo, hydrocarbon/butane monooxygenase; Pmo/Pxm, particulate methane monooxygenase. **b**, Cladogram detailing the affiliation of the *Nitrospira* sp.1 (green box) and sp.2 (red box) AmoA sequences within this sequence group. *Nitrospira* and *Crenothrix* sequences are depicted in bold. One representative sequence per study is shown for highly similar sequences; numbers in brackets indicate the number of sequences represented.

- Klotz, M. G. & Stein, L. Y. Nitrifier genomics and evolution of the nitrogen cycle. *FEMS Microbiol. Lett.* **278**, 146–156 (2008).
- Koch, H. *et al.* Expanded metabolic versatility of ubiquitous nitrite-oxidizing bacteria from the genus *Nitrospira*. *Proc. Natl Acad. Sci. USA* **112**, 11371–11376 (2015).
- Lupo, D. *et al.* The 1.3-Å resolution structure of *Nitrosomonas europaea* Rh50 and mechanistic implications for NH_3 transport by Rhesus family proteins. *Proc. Natl Acad. Sci. USA* **104**, 19303–19308 (2007).
- Daims, H. *et al.* Complete nitrification by *Nitrospira* bacteria. *Nature* <http://dx.doi.org/10.1038/nature16461> (2015).
- McTavish, H., Fuchs, J. A. & Hooper, A. B. Sequence of the gene coding for ammonia monooxygenase in *Nitrosomonas europaea*. *J. Bacteriol.* **175**, 2436–2444 (1993).
- Hyman, M. R. & Arp, D. J. $^{14}\text{C}_2\text{H}_2$ - and $^{14}\text{CO}_2$ -labeling studies of the de novo synthesis of polypeptides by *Nitrosomonas europaea* during recovery from acetylene and light inactivation of ammonia monooxygenase. *J. Biol. Chem.* **267**, 1534–1545 (1992).
- Taylor, A. E. *et al.* Use of aliphatic *n*-alkynes to discriminate soil nitrification activities of ammonia-oxidizing thaumarchaea and bacteria. *Appl. Environ. Microbiol.* **79**, 6544–6551 (2013).
- Ginestet, P., Audic, J.-M., Urbain, V. & Block, J.-C. Estimation of nitrifying bacterial activities by measuring oxygen uptake in the presence of the metabolic inhibitors allylthiourea and azide. *Appl. Environ. Microbiol.* **64**, 2266–2268 (1998).
- Zumft, W. G. Cell biology and molecular basis of denitrification. *Microbiol. Mol. Biol. Rev.* **61**, 533–616 (1997).
- Strous, M., Kuenen, J. G. & Jetten, M. S. Key physiology of anaerobic ammonium oxidation. *Appl. Environ. Microbiol.* **65**, 3248–3250 (1999).
- Alonso-Sáez, L. *et al.* Role for urea in nitrification by polar marine Archaea. *Proc. Natl Acad. Sci. USA* **109**, 17989–17994 (2012).
- Burton, S. A. Q. & Prosser, J. I. Autotrophic ammonia oxidation at low pH through urea hydrolysis. *Appl. Environ. Microbiol.* **67**, 2952–2957 (2001).
- Solomon, C., Collier, J., Berg, G. & Glibert, P. Role of urea in microbial metabolism in aquatic systems: a biochemical and molecular review. *Aquat. Microb. Ecol.* **59**, 67–88 (2010).
- Wagner, M., Nielsen, P. H., Loy, A., Nielsen, J. L. & Daims, H. Linking microbial community structure with function: fluorescence *in situ* hybridization-microautoradiography and isotope arrays. *Curr. Opin. Biotechnol.* **17**, 83–91 (2006).
- Daims, H., Nielsen, J. L., Nielsen, P. H., Schleifer, K. H. & Wagner, M. *In situ* characterization of *Nitrospira*-like nitrite-oxidizing bacteria active in wastewater treatment plants. *Appl. Environ. Microbiol.* **67**, 5273–5284 (2001).
- Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5–W9 (2008).
- Stoecker, K. *et al.* Cohn's *Crenothrix* is a filamentous methane oxidizer with an unusual methane monooxygenase. *Proc. Natl Acad. Sci. USA* **103**, 2363–2367 (2006).
- Luesken, F. A. *et al.* Diversity and enrichment of nitrite-dependent anaerobic methane oxidizing bacteria from wastewater sludge. *Appl. Microbiol. Biotechnol.* **92**, 845–854 (2011).

30. Glass, E. M., Wilkening, J., Wilke, A., Antonopoulos, D. & Meyer, F. Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harb. Protoc.* <http://dx.doi.org/10.1101/pdb.prot5368> (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank K. Stultiens, T. van Alen, J. Frank, P. Klaren, L. Pierson and L. Claessens-Joosten for technical assistance, T. Spanings for biofilter maintenance and C. Herbold for the ANI analysis. We are grateful for the use of the confocal microscope from the Microscopic Imaging Centre (MIC, Radboud UMC, Nijmegen) and would like to thank H. Croes and M. Willemse for technical assistance. The LABGeM team and the National Infrastructure “France Genomique” are acknowledged for support within the MicroScope annotation platform. We are thankful to C. Dupont, A. Santoro and M. Saito for consenting to our use of the *Nitrospira marina nxrA* sequences, which were produced by the US Department of Energy Joint Genome Institute. M.A.H.J.v.K was supported by the Technology Foundation STW (grant 13146), D.R.S. by the BE-Basic Foundation (grant fs7-002), M.A. and P.H.N. by the Danish Council for Independent Research (DFF 4005-00369),

M.S.M.J. by the European Research Council (ERC Advanced Grant projects anammox 232937 and Eco_MoM 339880) and the Dutch Ministry of Education, Culture and Science (Gravitation grant SIAM 024002002), B.K. and S.L. by the Netherlands Organization for Scientific Research (NWO VENI grants 863.11.003 and 863.14.019, respectively). The Radboud Excellence Initiative is acknowledged for support to S.L.

Author Contributions M.A.H.J.v.K and S.L. executed experiments and analysed data. D.R.S. and M.A. contributed to metagenomic data analyses. M.A. and P.H.N. performed sequencing, assembly and binning. M.A.H.J.v.K., H.J.M.O.d.C., B.K., M.S.M.J. and S.L. planned research. M.A.H.J.v.K., B.K. and S.L. wrote the paper. All authors discussed results and commented on the manuscript.

Author Information Metagenomic data is available in the European Nucleotide Archive (ENA) under accession numbers CZQA01000001–CZQA01000015 and CZPZ01000001–CZPZ01000036. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.L. (s.luecker@science.ru.nl).

METHODS

No statistical methods were used to predetermine sample size, the experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment

Enrichment and cultivation. A bioreactor was inoculated with biomass from a recirculation aquaculture system biofilter (3.5 l, obtained from the anoxic part of the trickling filter compartment) connected to an aquaculture system. The system accommodated common carp (*Cyprinus carpio*, approximately 3.5 kg total weight) and had a total volume of 900 l. The bioreactor (Applikon Biotechnology BV, Schiedam, The Netherlands) consisted of stainless steel and glass, had a 7 l working volume, was equipped with pH and dissolved oxygen sensors (Applikon Dependable Instruments BV, Applisens, Schiedam, The Netherlands) and connected to an ADI1030 biocontroller (Applikon Biotechnology BV, Schiedam, The Netherlands). It was operated as a sequencing batch reactor (SBR) with 12 h or 24 h cycles. In the first 5 months, the reactor was operated with a 24 h cycle that consisted of 23 h 15 min filling, 15 min settling (no stirring) and 30 min removal of the supernatant. Afterwards, in 12 h cycles, each filling cycle consisted of 11 h 15 min, followed by 15 min settling and 30 min removal of the supernatant. During every filling period, the reactor was supplied with 600 ml of medium (0.83 ml min^{-1}). The reactor and the medium were flushed constantly with Ar/CO_2 (95%/5% v/v, 10 ml min^{-1}). The temperature was kept at $23 \pm 1^\circ\text{C}$ with a heating blanket and pH was maintained at 6.99 ± 0.1 using a 1 M KHCO_3 solution. The reactor was stirred at 200 r.p.m. Medium was prepared using aquaculture water taken from the recirculation aquaculture system biofilter. This water contained $300\text{--}1,848 \mu\text{M NO}_3^-$, $0\text{--}29 \mu\text{M NO}_2^-$ and $0\text{--}75 \mu\text{M NH}_4^+$. The water was filter-sterilized (polysulfone filter HF80S, Fresenius Medical Care, Bad Homburg, Germany) and supplemented with $100\text{--}500 \mu\text{M NH}_4^+$, $100\text{--}450 \mu\text{M NO}_2^-$ and $500 \mu\text{M NO}_3^-$.

DNA extraction and genome sequencing. DNA was extracted using the PowerSoil DNA isolation kit (MoBio, Carlsbad, CA) or a CTAB-based extraction method³¹. 1 μg of DNA was used to prepare paired-end sequencing libraries using the TruSeq PCR-free kits (Illumina, San Diego, CA, USA) following the manufacturers recommendation except that the 550 bp protocol was used with 1 μg of input DNA. Mate-pair libraries were prepared using the Nextera Mate-pair kit (Illumina) using the gel-free approach. The prepared libraries were sequenced using an Illumina MiSeq with MiSeq Reagent Kit v3 ($2 \times 301 \text{ bp}$; Illumina).

Bioinformatics. Data generation and binning of metagenome scaffolds to individual genome bins was conducted as described in the mmgenome workflow³² which builds on the multi-metagenome principles⁵. Paired-end Illumina reads in FASTQ format were imported to CLC Genomics Workbench v.8.0 (CLC Bio, Aarhus, Denmark) and trimmed using a minimum phred score of 20, a minimum length of 50 bp, allowing no ambiguous nucleotides and trimming off Illumina sequencing adaptors. Mate-pair reads in FASTQ format were trimmed using NextClip³³ and only reads in class A were used for assembly. Passing reads were co-assembled using CLCs *de novo* assembly algorithm, using a kmer of 63 and a minimum scaffold length of 1 kbp. The trimmed metagenome reads were afterwards independently mapped to the assembled scaffolds using CLCs 'map reads to reference' algorithm, with a minimum similarity of 95% over 80% of the read length.

Open reading frames were predicted in the assembled scaffolds using the metagenome version of Prodigal³⁴. A set of 107 HMMs of essential single-copy genes³⁵ were searched against the predicted open reading frames using HMMER3³⁶ with default settings, except for the use of the trusted cut-off (-cut_tc). Identified proteins were taxonomically classified using BLASTP against the RefSeq (version 52) protein database with a maximum *e*-value cut-off of 10^{-5} . MEGAN³⁷ was used to extract class level taxonomic assignments from the BLAST .xml output file. The script network.pl was used to extract paired-end read connections between scaffolds using a SAM file of the read mappings to the metagenome.

Individual genome bins were extracted using the multi-metagenome principles⁵ and refined using tetranucleotide frequencies, as implemented in the mmgenome R package³². The script extract.fastq.reassembly.pl was used to extract paired-end reads from the binned scaffolds, which were used for re-assembly using SPAdes 3.5.0³⁸. Paired-end and mate-pair connections were used to manually refine the extracted genome bins. For all genomes quality was assessed using coverage plots through the mmgenome R package and by the use of QUAST³⁹ and CheckM⁴⁰ (see Supplementary Table 2 for CheckM counts of single-copy genes). Manual inspection of potential misassemblies was done using Circos⁴¹ as described³². In addition, key regions were manually inspected in CLC Genomics Workbench.

The *Nitrospira* draft genomes were integrated into the MicroScope annotation platform⁴². The automatic annotation of genes in key metabolic pathways was manually refined using the respective tools in MaGe⁴³ as described previously⁷. Genomic pairwise average nucleotide identity values were calculated using BLAST (ANi) in JSpecies⁶.

Absence of canonical bacterial or archaeal *amoA* sequences in the metagenome data was confirmed by searching a set of reference sequences against a BLAST database containing all trimmed metagenome reads.

Code availability. The Rmarkdown files used for extracting the genome bins are available for download³².

Activity assays. For activity assays, the reactor was supplied with medium containing labelled ammonium ($^{15}\text{NH}_4^+$). The medium flow was kept at normal operating rate (0.83 ml min^{-1}) and the biomass was stirred continuously. Isotopic composition of the nitrogen gas produced was analysed using gas chromatography (Agilent 6890 equipped with a Porapak Q column at 80°C and a TCD detector at 300°C ; Agilent Technologies, Santa Clara, CA, USA) combined with mass spectrometry (Agilent 5975c, quadrupole inert MS).

For batch assays, 150 ml biomass was taken from the reactor and harvested by centrifugation (300g, 10 min). Flocs were disrupted by resuspending the biomass in 1.5 ml mineral medium⁴⁴, followed by rigorous horizontal shaking in the presence of a 0.75 inch glass sphere for 10 min. Subsequently, biomass was washed twice in mineral medium and resuspended in 150 ml mineral medium containing no N-source. 12 ml biomass per incubation was transferred to 30 ml serum bottles and ammonium, nitrite or urea was added ($200 \mu\text{M}$ final concentration). To test for anammox activity and denitrification $^{15}\text{NH}_4^+$ was used and the headspace analysed for labelled dinitrogen gas production as described above. For inhibition experiments ATU was added to a final concentration of $100 \mu\text{M}$ and biomass was preincubated for 10 min before substrate addition. Bottles were sealed with rubber stoppers and 10 ml air was added to the headspace to ensure slight overpressure. Incubations were performed at room temperature in the dark with mild agitation (50 rpm). At each time point, 0.5 ml sample was taken and stored at -20°C for further analysis.

Analytical methods. Ammonium was determined colorimetrically using a modified orthophthalaldehyde assay⁴⁵ (detection limit $10 \mu\text{M}$) and nitrite ($\geq 5 \mu\text{M}$) by the sulfanilamide reaction⁴⁶. Nitrate ($\geq 1 \mu\text{M}$) was measured by converting it into nitric oxide at 95°C using a saturated solution of VCl_3 in HCl. Nitric oxide was then measured using a nitric oxide analyser (NOA280i, GE Analytical Instruments, Manchester, UK). To determine the total organic carbon (TOC) concentration of the medium, medium was first acidified to remove inorganic carbon. After $6.5 \times$ dilution with ultrapure water, samples were measured using a TOC-L CPH/CPN analyser (Shimadzu, Duisburg, Germany).

Fluorescence *in situ* hybridization (FISH). For FISH analysis, samples from the reactor were fixed with 4% (v/v) paraformaldehyde (PFA), followed by hybridization with fluorescently labelled oligonucleotides as described elsewhere⁴⁷. FISH probes used in this study (Extended Data Table 2) were 5' labelled with the dyes FLUOS (5(6)-carboxyfluorescein-N-hydroxysuccinimide ester), Cy3 or Cy5 (Thermo Electron Corporation, Ulm, Germany). After hybridization, slides were air-dried and embedded in Vectashield (Vector Laboratories Inc., Burlingame, CA). Probe-conferred fluorescence was recorded on an Zeiss Axioplan 2 (Carl Zeiss AG, Oberkochen, Germany) equipped with a HBO 100 light source and specific filter sets for the detection of FLUOS, Cy3 and Cy5, a Leica TCS SP2 AOBs (Leica Microsystems, Wetzlar, Germany) or a Zeiss LSM510 META (Carl Zeiss AG) confocal laser scanning microscope (CLSM), both equipped with one argon ion (450–514 nm) and two helium neon lasers (543 and 633 nm). Images were recorded with $63 \times$ glycerol or oil immersion objectives at a resolution of $1,024 \times 1,024$ pixels and 8-bit depth.

For quantifying relative biovolume fractions, PFA-fixed reactor biomass was hybridized to probes Ntsa662, Amx820 and EUB338mix (Extended Data Table 2) as described above. Subsequently, 45 image pairs were recorded at random fields of view using the Leica TCS SP2 AOBs CLSM. The images were imported into the image analysis software daime⁴⁸ and evaluated as described elsewhere⁴⁹.

AMO-labelling. Washed and disrupted (see above) biomass was incubated for 30 min at room temperature with freshly prepared fluorescein thiocarbonyl-propargylamine (FTCP, synthesized as described elsewhere¹⁶). After incubation, cells were harvested, washed, PFA-fixed and hybridized to specific FISH probes as described above.

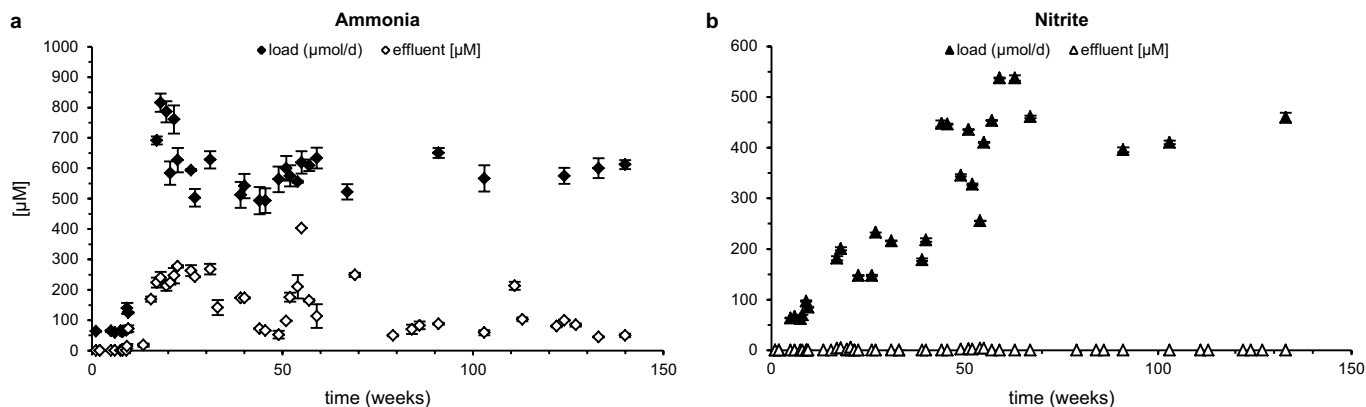
FISH combined with microautoradiography (FISH-MAR). FISH-MAR experiments were performed as described before⁵⁰. 150 ml biomass was taken from the reactor and flocs were disrupted as described above. After harvesting and washing, the biomass was resuspended in mineral medium⁴⁴ and transferred to serum bottles. Ammonium or nitrite was added to a final concentration of $500 \mu\text{M}$. As controls, incubations with ammonium and ATU ($100 \mu\text{M}$), without nitrogen source and a dead control (PFA-fixed biomass) were performed. $10 \mu\text{Ci}$ [^{14}C]-labelled bicarbonate were added to all samples, bottles were sealed with rubber stoppers and incubated at room temperature in the dark for 18 h. After incubation, the biomass was harvested by centrifugation (20,000g, 10 min), PFA-fixed and FISH was performed on coverslips as described above. Hybridized samples were dipped in preheated (48°C) and diluted (1:1 with deionised water) film emulsion (Ilford

Nuclear Emulsion K5, Harman Technology, UK). After overnight drying at room temperature, samples were exposed for 6 days at 4 °C and developed in Kodak D19 developer as described before⁵⁰. Images were recorded on a Zeiss LSM510 META CLSM as detailed above. To correct for the different levels of unspecific silver grain deposition in the incubations, the degree of silver grain formation in areas without biomass was compared to the amount of silver grains above biomass flocs. Only cell clusters which showed grain deposition clearly above background level were considered positive.

Phylogenetic analyses. 16S rRNA sequences with nucleotide identities $\geq 98\%$ and *amoA* sequences with identities $\geq 70\%$, to the respective sequences of *Nitrospira* sp.1 or sp.2 were identified in the NCBI nr database by BLAST²⁷. 16S rRNA sequences were imported into the SILVA⁵¹ small subunit ribosomal RNA database release 119, *amoA* sequences in a custom-made database containing a reference set of *amoA* and *pmoA* sequences. *nrrA* sequences were imported in a custom-made database containing all published sequences from *Nitrospira*, *Nitrospina* and anammox organisms. Sequence alignments for all data sets were generated and manually refined using ARB 5.5⁵². Bayesian inference trees were calculated using MrBayes 3.2.3⁵³ until a standard deviation < 0.01 was reached. For 16S rRNA analyses the GTR substitution model and a 50% conservation filter resulting in 1463 valid alignment positions were used. *amoA* genes were translated into their amino acid sequence and a 10% conservation filter resulting in 264 alignment positions in combination with the WAG substitution model were used for tree calculation. *nrrA* trees were calculated from nucleic acid sequences with the GTR substitution model and without conservation filter, resulting in 2,660 distinct alignment patterns. For all trees 50% majority rule consensus trees are shown.

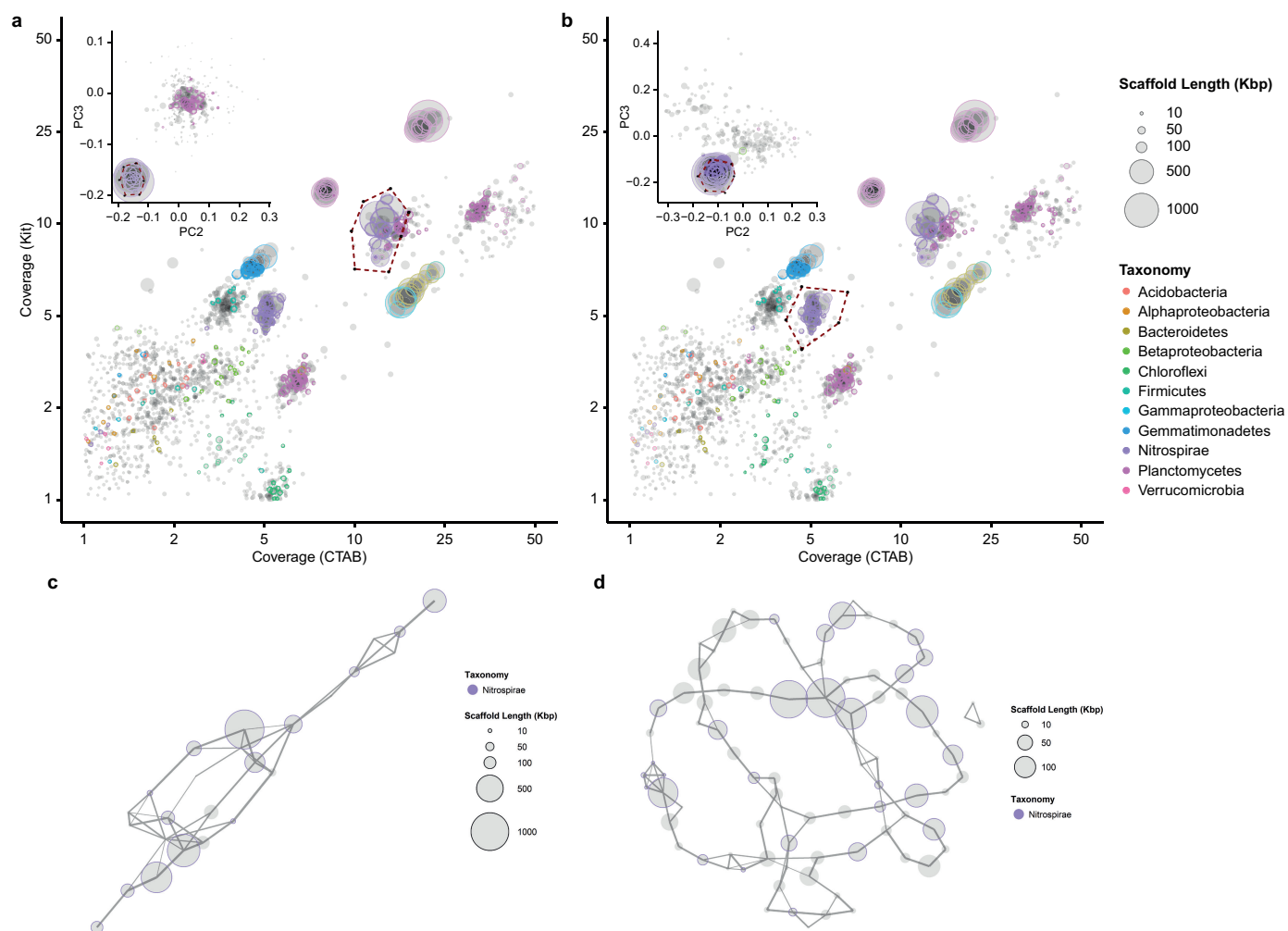
Database mining. All 7,306 public shotgun metagenomes and metatranscriptomes available in MG-RAST⁵⁴ were searched for the presence of the diagnostic *amoA* gene. Data sets were downloaded and searched against a small set of characteristic *amoA* sequences using DIAMOND⁵⁵ with the default settings. The resulting 44,993 hits were filtered using a BLAST score ratio⁵⁶ of the initial alignment score versus the alignment score against the NCBI nr.

31. Zhou, J., Bruns, M. A. & Tiedje, J. M. DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **62**, 316–322 (1996).
32. Albertsen, M. mmgenome: tools for extracting individual genomes from metagenomes <http://madsalbertsen.github.io/mmgenome/> (2015).
33. Leggett, R. M., Clavijo, B. J., Clissold, L., Clark, M. D. & Caccamo, M. NextClip: an analysis and read preparation tool for Nextera Long Mate Pair libraries. *Bioinformatics* **30**, 566–568 (2014).
34. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
35. Dupont, C. L. et al. Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J.* **6**, 1186–1199 (2012).
36. Eddy, S. R., Wheeler, T. J. & the HMMER development team. HMMER: biosequence analysis using profile hidden Markov models. <http://hmmerr.janelia.org/> (2015).
37. Huson, D. H., Mitra, S., Ruscheweyh, H. J., Weber, N. & Schuster, S. C. Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* **21**, 1552–1560 (2011).
38. Pribelski, A. D. et al. ExSPAnDer: a universal repeat resolver for DNA fragment assembly. *Bioinformatics* **30**, i293–i301 (2014).
39. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
40. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
41. Krzywinski, M. et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
42. Vallenet, D. et al. MicroScope—an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucleic Acids Res.* **41**, D636–D647 (2013).
43. Vallenet, D. et al. MaGe: a microbial genome annotation system supported by synteny results. *Nucleic Acids Res.* **34**, 53–65 (2006).
44. Spieck, E. & Lipski, A. in *Methods in Enzymology* Vol. 486 (ed Martin, G. K.) 109–130 (Academic Press, 2011).
45. Taylor, S., Ninjoor, V., Dowd, D. M. & Tappel, A. L. Cathepsin B2 measurement by sensitive fluorometric ammonia analysis. *Anal. Biochem.* **60**, 153–162 (1974).
46. Griess, P. Bemerkungen zu der Abhandlung der HH. Weselsky und Benedikt „Ueber einige Azoverbindungen“. *Ber. Dtsch. Chem. Ges.* **12**, 426–428 (1879).
47. Daims, H., Stoecker, K. & Wagner, M. in *Molecular Microbial Ecology* (eds Osborn, A. M. & Smith, C. J.) Ch. 9, 213–239 (Taylor & Francis, 2005).
48. Daims, H., Lückner, S. & Wagner, M. *daime*, a novel image analysis program for microbial ecology and biofilm research. *Environ. Microbiol.* **8**, 200–213 (2006).
49. Daims, H. & Wagner, M. Quantification of uncultured microorganisms by fluorescence microscopy and digital image analysis. *Appl. Microbiol. Biotechnol.* **75**, 237–248 (2007).
50. Lee, N. et al. Combination of fluorescent *in situ* hybridization and microautoradiography—a new tool for structure-function analyses in microbial ecology. *Appl. Environ. Microbiol.* **65**, 1289–1297 (1999).
51. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
52. Ludwig, W. et al. ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (2004).
53. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
54. Meyer, F. et al. The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**, 386 (2008).
55. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* **12**, 59–60 (2015).
56. Rasko, D. A., Myers, G. S. & Ravel, J. Visualization of comparative genomic analyses by BLAST score ratio. *BMC Bioinformatics* **6**, 2 (2005).
57. Schmid, M. et al. Molecular evidence for genus level diversity of bacteria capable of catalyzing anaerobic ammonium oxidation. *Syst. Appl. Microbiol.* **23**, 93–106 (2000).
58. Stahl, D. A. & Amann, R. in *Nucleic Acid Techniques in Bacterial Systematics* (eds Stackebrandt, E. & Goodfellow, M.) (Wiley, 1991).
59. Amann, R. I. et al. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl. Environ. Microbiol.* **56**, 1919–1925 (1990).
60. Daims, H., Brühl, A., Amann, R., Schleifer, K. H. & Wagner, M. The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. *Syst. Appl. Microbiol.* **22**, 434–444 (1999).
61. Wagner, M., Rath, G., Amann, R., Koops, H.-P. & Schleifer, K.-H. *In situ* identification of ammonia-oxidizing bacteria. *Syst. Appl. Microbiol.* **18**, 251–264 (1995).
62. Juretschko, S. et al. Combined molecular and conventional analyses of nitrifying bacterium diversity in activated sludge: *Nitrosococcus mobilis* and *Nitrospira*-like bacteria as dominant populations. *Appl. Environ. Microbiol.* **64**, 3042–3051 (1998).
63. Mobarry, B. K., Wagner, M., Urbain, V., Rittmann, B. E. & Stahl, D. A. Phylogenetic probes for analyzing abundance and spatial organization of nitrifying bacteria. *Appl. Environ. Microbiol.* **62**, 2156–2162 (1996).
64. Alm, E. W., Oerther, D. B., Larsen, N., Stahl, D. A. & Raskin, L. The oligonucleotide probe database. *Appl. Environ. Microbiol.* **62**, 3557–3559 (1996).



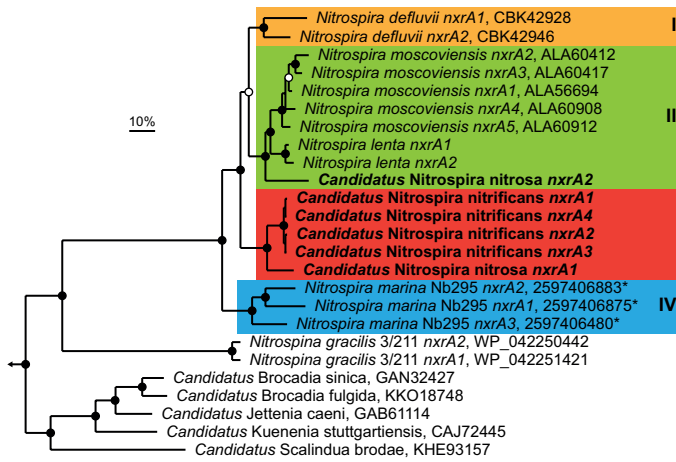
Extended Data Figure 1 | Ammonium and nitrite conversion by the enrichment culture. **a, b,** Inorganic nitrogen load of the enrichment culture per 24 h cycle (filled symbols) and effluent concentrations (open symbols) for ammonium (**a**, diamonds) and nitrite (**b**, triangles). Effluent nitrite concentrations were below the detection limit ($<5\mu\text{M}$) at all time points. Data points represent the mean of three technical

replicates, error bars the standard deviations of these triplicates. Nitrate concentration in the medium varied between 0.5 and 2.0 mM and total organic carbon (TOC) content between 1.30 and 1.44 p.p.m., which was due to medium preparation with water obtained directly from the recirculation aquaculture system.



Extended Data Figure 2 | Metagenome binning. **a, b**, Extraction of the *Nitrospira* sp.1 (**a**) and sp.2 (**b**) genome sequences from the metagenome using differential coverage binning. Each circle represents a metagenomic scaffold, with size proportional to scaffold length; the plots contain a total of 47,584 scaffolds. The inlay of each figure shows the secondary binning based on tetranucleotide frequencies, with a total of 331 (**a**) and 281 (**b**)

scaffolds included. Taxonomic classification is indicated by colour; a total of 3,158 essential marker genes were detected. The extracted bins are enclosed by a dashed line. **c, d**, Genome contaminations were excluded by generating linkage maps of the final bins of sp.1 (**c**, 25 scaffolds) and sp.2 (**d**, 86 scaffolds) using mate-pair sequencing data.

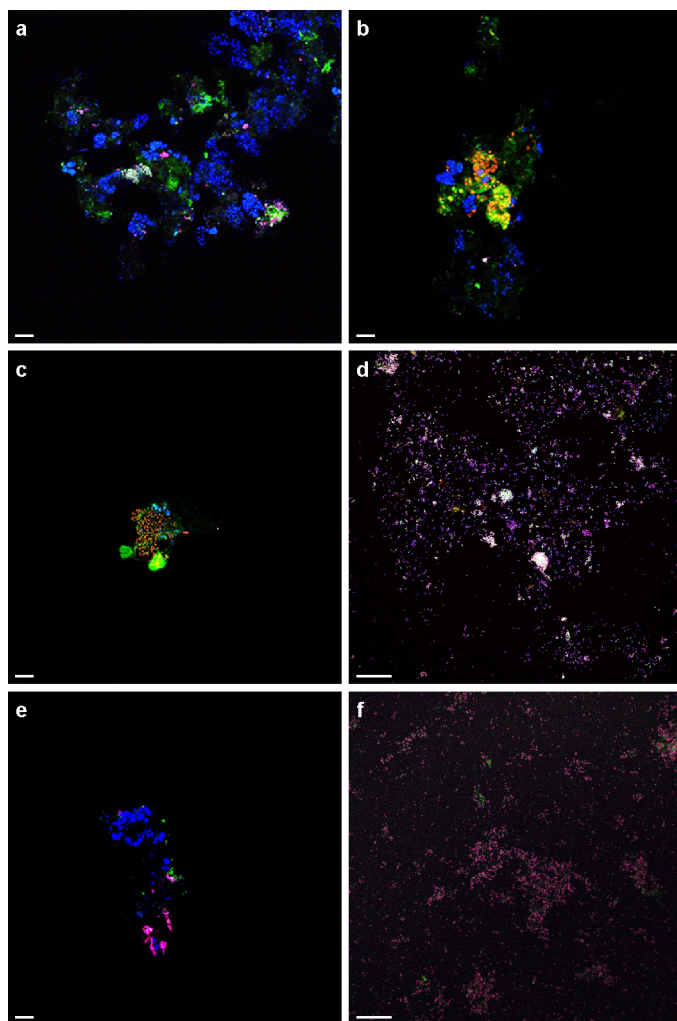


Extended Data Figure 3 | Phylogenetic analysis of NXR. Bayesian inference tree (s.d. = 0.0099) showing the affiliation of the *Nitrospira* sp.1 and sp.2 *nrrA* sequences in comparison to other genome-sequenced *Nitrospira*, *Nitrospina* and anammox bacteria. Posterior probabilities $\geq 70\%$ and $\geq 90\%$ are indicated by open and filled circles, respectively. NCBI protein accession numbers for all publicly available sequences are indicated, numbers with an asterisk are IMG gene IDs. The described *Nitrospira* sublineages are indicated by coloured boxes and roman numbers. The scale bar represents 10% sequence divergence. Note the different affiliation of the 'Candidatus N. nitrosa' (sp.1) *nrrA* sequences. The tree contains 25 sequences from 12 species, belonging to 3 different phyla. Sequences from closely related bacterial putative nitrate reductases were used as outgroup ($n = 4$); the outgroup position is indicated by the arrow.



Extended Data Figure 4 | 16S rRNA-based phylogenetic analysis. Bayesian inference tree (s.d. = 0.0098) showing the affiliation of the *Nitrospira* sp.1 and sp.2 16S rRNA sequences within *Nitrospira* sublineage II. Posterior probabilities $\geq 70\%$ and $\geq 90\%$ are indicated by open and filled circles, respectively. The strongly supported sequence group containing the novel *Nitrospira* spp. catalysing complete nitrification is shaded in grey, the two subgroups containing *Nitrospira* sp.1 and sp.2 (in bold) are highlighted by green and red boxes, respectively.

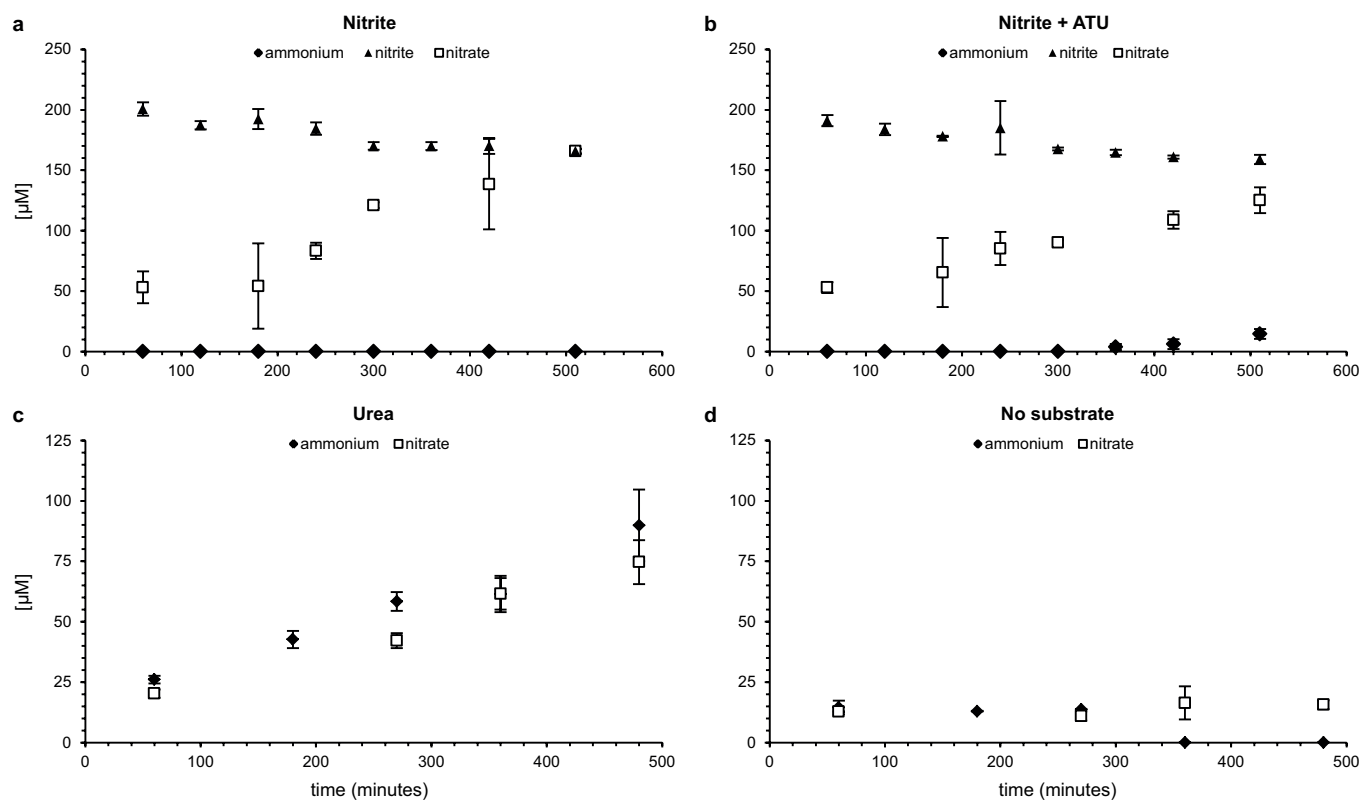
N. moscoviensis is depicted in bold for comparison. The curly bracket indicates the target group of the newly designed FISH probe Ntspa476 (see Extended Data Table 2). Scale bar indicates 10% sequence divergence. The tree contains a total of 181 sequences; the size of sequence groups is indicated in brackets. Sequences from members of *Nitrospira* sublineages I and IV were used as outgroup ($n = 24$); the outgroup position is indicated by the arrow.



Extended Data Figure 5 | Control experiments of AMO-labelling.

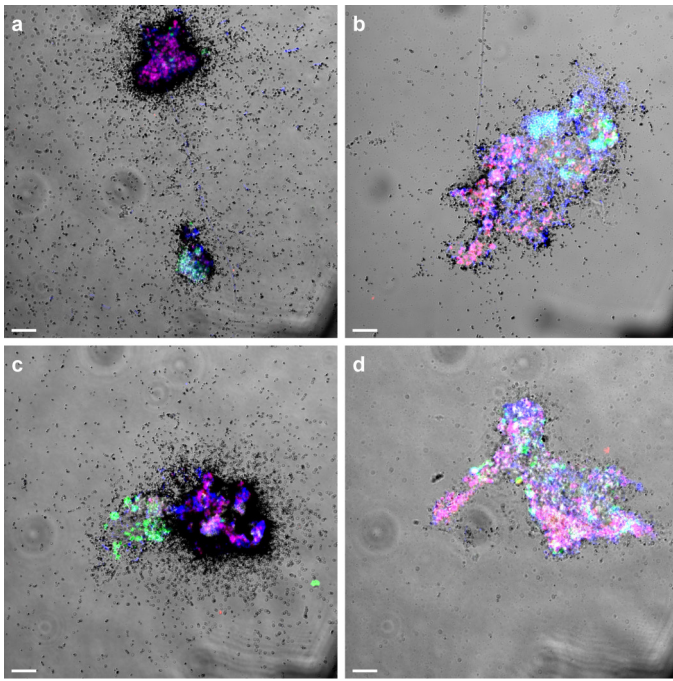
a, Cells incubated with the fluorescent dye FTCP (green) were stained by FISH using probes specific for *Nitrospira* (Ntspa662, red) and all bacteria (EUB338mix, blue). A small cell cluster was stained by FTCP and targeted by both probes (resulting in a white overlay signal), while all other bacteria (in blue) were not or only slightly stained by FTCP. The green signal is due to autofluorescence and unspecific FTCP binding to the floc matrix.

b, Anammox cells (Amx820, blue) showed minor staining by FTCP (green), but to a much lesser degree than *Nitrospira* (Ntspa662, red; yellow overlay). **c** and **d**, Positive controls: ammonium oxidizing bacteria (c, Nso1225 and Nso190, red) in an aerobic enrichment culture and a *Nitrosomonas europaea* pure culture (d, NEU, red, and EUB338mix, blue) were stained by FTCP (resulting in yellow and white overlays, respectively). **e** and **f**, Negative controls: canonical *Nitrospira* in an aerobic enrichment culture (e, Ntspa662, blue) and a *Nitrospira moscoviensis* pure culture (f, Ntspa662, red, and EUB338mix, blue; magenta overlay) did not show any labelling with FTCP (green). The two bright green structures in (c) and the bright pink signal in (e) are due to autofluorescence. Images are representative of two (a and b) or one (c to f) individual experiments, with three technical replicates each. Scale bars in all panels represent 10 μm .



Extended Data Figure 6 | Batch incubations with nitrite, urea and without substrate. **a, b**, Nitrite (triangles) oxidation by the enrichment culture to nitrate (squares) in the absence (**a**) and in the presence (**b**) of ATU. The ammonia (diamonds) in **b** presumably stems from biomass decay and is not oxidized owing to ATU inhibition. **c**, Urea conversion to ammonium (diamonds) and subsequent oxidation to nitrate (squares). **d**, No-substrate control; minor amounts of ammonium (diamonds)

presumably stem from mineralisation of degrading biomass, leading subsequently to nitrate (squares) formation. Symbols in all plots represent averages of three independent incubations; ammonium was determined in single measurements, nitrite and nitrate in duplicate (**a** and **b**) or triplicate (**c** and **d**). Error bars represent standard deviations of three biological replicates.



Extended Data Figure 7 | Ammonium and nitrite-dependent CO₂ fixation shown by FISH-MAR. a–d, FISH with probes for all bacteria (EUB338mix, blue), and probes specific for *Nitrospira* (Ntspa662, red; resulting in magenta) and anammox bacteria (Amx820, green; resulting in cyan). **a**, Ammonia-dependent carbon fixation. Only *Nitrospira* cells were active, as indicated by silver grain deposition. Note the inactive anammox cells on the left side of the smaller floc, co-localizing with highly active *Nitrospira* cells on the right side of the same floc. **b**, Inhibition of ammonia-dependent carbon fixation by ATU. **c**, Nitrite-dependent carbon fixation. Only *Nitrospira* cells incorporated ¹⁴CO₂. **d**, No-substrate control. Images are representative of two individual experiments, with two technical replicates each. Scale bars in all panels represent 10 μm.

Extended Data Table 1 | General genomic characteristics of *Nitrospira* sp.1 and sp.2

Bin	Ca. N. nitrosa (sp.1)		Ca. N. nitrificans (sp.2)	
	initial	final	initial	final
Genome size (bp)	4413075	4422398	4088547	4117083
Contigs	25	15	86	36
Largest contig (bp)	1073143	1804237	335390	475968
N50	659693	727365	103850	174194
# Ns per 100 Kbp	355	0	420	0
Completeness [*]	99% (97%)	>99% (97%)	>95% (97%)	>95% (97%)
Contamination [*]	0% (2.3%)	0% (2.3%)	<1% (2.8%)	<1% (2.7%)
Coverage (CTAB) [†]	~‡	13.0	~‡	4.9
Coverage (Kit) [†]	~‡	10.0	~‡	5.0
Average G+C content	~‡	54.8	~‡	56.6
Number of coding sequences (CDS)	~‡	4309	~‡	4502
rRNA operons	~‡	1	~‡	1
tRNAs	~‡	46	~‡	43

^{*}Values are based on evaluation of the binning plots and manual inspection; numbers in brackets are based on CheckM⁴⁰

[†]For details on DNA extraction see Methods section.

[‡]These values were only determined for the final genomic bins.

Extended Data Table 2 | FISH probe specifications

Probe name	Probe full name*	Sequence (5'-3')	Binding position†	FA%‡	Specificity	Ref.
Amx820	S-*-Amx-0820-a-A-22	AAA ACC CCT CTA CTT AGT GCC C	820 - 841	40	Genera <i>Brocadia</i> , <i>Kuenenia</i>	57
Arch915	S-D-Arch-0915-a-A-20	GTG CTC CCC CGC CAA TTC CT	915 - 934	nd§	Domain <i>Archaea</i>	58
Eub338	S-D-Bact-0338-a-A-18	GCT GCC TCC CGT AGG AGT	338 - 355	0 - 50	Domain <i>Bacteria</i>	59
Eub338	S-*-Bact-0338-b-A-18	GCA GCC ACC CGT AGG TGT	338 - 355	0 - 50	Order <i>Planctomycetales</i>	60
Eub338	S-*-Bact-0338-c-A-18	GCT GCC ACC CGT AGG TGT	338 - 355	0 - 50	Order <i>Verrucomicrobiales</i>	60
NEU	S-*-Nsm-0651-a-A-18	CCC CTC TGC TGC ACT CTA	653 - 670	40	<i>Nitrosomonas</i> spp.	61
cNEU	-	TTC CAT CCC CCT CTG CCG	659 - 676	-	Competitor to NEU	61
NmV	S-S-Nmob-0174-a-A-18	TCC TCA GAG ACT ACG CGG	174 - 191	35	<i>Nitrosococcus mobilis</i> lineage	62
Nso190	S-F-bAOB-0189-a-A-19	CGA TCC CCT GCT TTT CTC C	189 - 207	55	Betaproteobacterial AOB	63
Nso1225	S-F-bAOB-1224-a-A-20	CGC CAT TGT ATT ACG TGT GA	1224 - 1243	35	Betaproteobacterial AOB	63
Ntspa662	S-G-Ntspa-662-a-A-18	GGA ATT CCG CGC TCC TCT	662 - 679	35	Genus <i>Nitrospirae</i>	26
cNtspa662	-	GGA ATT CCG CTC TCC TCT	662 - 679	-	Competitor to Ntspa662	26
Ntspa712	S-*-Ntspa-712-a-A-21	CGC CTT CGC CAC CGG CCT TCC	712 - 732	35	Phylum <i>Nitrospirae</i>	26
cNtspa712	-	CGC CTT CGC CAC CGG TGT TCC	712 - 732	-	Competitor to Ntspa712	26
Ntspa476	S-*-Ntspa-0476-a-A-22	CTG CAG GTA CCG TCC GAA	476 - 494	20	<i>Ca. N. nitrosa</i> , <i>Ca. N. nitrificans</i>	This study
cNtspa476	-	CTG GAG GTA CCG TCC GAA	476 - 494	-	Competitor to Ntspa476	This study

*Probe nomenclature according to Alm *et al.*⁶⁴†Probe binding position according to *Escherichia coli* 16S rRNA gene numbering.

‡Percent formamide (v/v) added to the hybridization buffer for optimal hybridization stringency.

§Not determined.

^{||}Probes were used in an equimolar mixture (EUB338mix) to detect all *Bacteria*.*Probe targets *N. mobilis*, which is affiliated with the betaproteobacterial *Nitrosomonas* lineage and not the gammaproteobacterial genus *Nitrosococcus*.

References 57–64 are cited in this table.

Extended Data Table 3 | Metagenome screening for *Nitrospira*-like *amoA* sequences

Source	Geographical location	Number of hits	Total reads [†]	Project name	Dataset ID [‡]
Metagenome projects					
River sediment	Tongue river, Montana, USA	1327	556,961,375	Tongue_all_2011	4481956-57; 63-72; 74-86
Soil	Houston, Texas, USA	367	321,988,632	Metagenomic investigation for a ethanol-blended fuel spill	4519753-58; 60-64, 67-76
Prairie soil	Auburn, Illinois, USA	119	1,075,325,181	ISA-SMC-2011	4502539-2541; 2543; 2923-2924; 2926; 2928; 2930; 2932-33; 2935
Soil	Ha Noi, Vietnam	94	246,030,284	Rice field	4626743-47; 53-54
Garden soil	Xiamen, Fujian, China	80	46,831,964	¹³ C labeling Soil Metagenome	4635904-5
Air	Beijing, China	68	978,592,643	Beijing PM2.5 and MP10 Pollutants	4516402-6403; 6455; 6459; 6637; 6651; 6802-6803; 6910-6911; 6952; 7064
Agricultural soil	Amazonia, Brazil	63	254,067,071	Amazon Soil metagenome 2_mendes	4497370-371; 376; 391-393; 395-396; 407-409; 411-412
Marine sediment	Gulf of Mexico, USA	45	2,425,926,864	BP_Sediments	4510162-66; 68-69; 71; 73-74
Marine sediment	Plum Island, Massachusetts, USA	33	38,370,475	IGERT Reverse Ecology 2011-2013	4519628; 19632; 19636; 20031
Activated sludge	Stanley wwtp, Hong Kong	26	16,663,946	Stanley wwtp activated sludge sample	4467420
Soil	Danum, Malaysia	24	43,344,688	Effect of logging on soil microbial community in tropics	4582264-267; 270; 798; 802-803; 805
Agricultural soil	Richmond, Indiana, USA	23	70,731,826	EarlhamMetagenomes2012	4508937-38; 40
wwtp sludge	Malaysia	23	40,000,000	UTM waste water treatment plant project A	4544292-4293; 4301; 4307; 5190; 6367-6368; 6370; 6373; 6375
Activated sludge	Switzerland	20	9,455,087	Swiss wwtp metatranscriptomics	4491800
Soil	Cologne, Germany	20	46,128,675	Barley	4529836; 30504
Alkaline travertine water	Voltri Massif, Liguria, Italy	19	42,594,481	Microbial Biogeography of Serpentinities	4537864-69
Soil	Iowa, USA	16	790,560,095	GP corn unassembled	4539519; 21; 23; 28
Sports facility soil	Norman, Oklahoma, USA	15	10,247,092	Natural products	4573678; 83
River water	Minnesota, USA	14	60,806,478	M3P 2012	4534334-35; 45-47
Ochard soil	Haifa, Israël	13	27,265,311	Revital_aft_qc	4631721; 24
Freshwater sediment	Rifle, Colorado, USA	8	236,916,472	Subsurface Rifle	4465820; 4465822
Rizosphere	Golm, Germany	8	32,897,323	Barley_Rhizomicrobiomics_test_B_PE	4524591; 96
Coral reef	Xisha island, China	8	125,160,089	S_TS_MG	4580696-698; 702
Soil	Basque Country, Spain	6	3,293,845	Metal_soil	4510865
Mine soil	Coto Txomin, Spain	6	196,440	Pb-Zn-Mine	4580863; 73
River biofilm	West Virginia, USA	6	3,487,276	MTR_GeMS_DNA	4589540-1
Marine sediment	Santa Barbara, California, USA	5	96,123,985	Scott_Nitro	4537093
Cave microbial mat	Weebubbe cave, Eucla, Australia	4	475,608	Weebubbe Cave Slime Curtain Metagenome	4448052
Groundwater	Tulum, Quintana Roo, Mexico	4	59,482,508	Yucatan Groundwater	4536382-3
Grassland soil	Bethel, Minnesota, USA	4	71,162,444	CedarCreek_minsoil_june2013	4541645
Soil	Amazonia, Brazil	3	23,648,292	Amazon Soil metagenome 1	4493652
Freshwater microbial mat	Hot creek, Colorado, USA	3	6,877,377	International geobiology course 02014 PreTrip	4549766
River sediment	Athabasca, Alberta, Canada	2	2,524,335	Athabasca-biofilms	4482887
Metatranscriptome projects					
River microbial mat	West Virginia, USA	523	174,983,655	MTR_GeMS_RNA	4597881-86
Oil contaminated soil	Varenes, Quebec, Canada	164	234,156,703	GenoRem_GH_MT	4512573; 576-580; 586; 590; 592; 608
Soil	Kalamazoo, Michigan, USA	28	205,252,966	Miscanthus Metatranscriptome	4554103
Marine sediment	Gulf of Mexico, USA	9	152,742,090	MG-Core_Metat_Merged	4508038; 41; 53
Paddy soil	Jiangdu, China	6	52,988,024	paddy soil	4553284-5

*Number of sequences affiliated with the novel *AmoA*/unusual *PmoA* sequence group.[†]Total number of metagenomic reads in the respective MG-RAST project.[‡]For retrieving these datasets from MG-RAST '3' must be added to the respective dataset ID.

Interleukin-22 promotes intestinal-stem-cell-mediated epithelial regeneration

Caroline A. Lindemans^{1,2*}, Marco Calafiore^{1*}, Anna M. Mertelsmann^{1*}, Margaret H. O'Connor^{1*}, Jarrod A. Dudakov^{3,4}, Robert R. Jenq^{1,5}, Enrico Velardi³, Lauren F. Young³, Odette M. Smith³, Gillian Lawrence¹, Juliet A. Ivanov¹, Ya-Yuan Fu¹, Shuichiro Takashima¹, Guoqiang Hua^{6,7}, Maria L. Martin⁷, Kevin P. O'Rourke⁸, Yuan-Hung Lo⁹, Michal Mokry², Monica Romera-Hernandez¹⁰, Tom Cupedo¹⁰, Lukas E. Dow⁵, Edward E. Nieuwenhuis², Noah F. Shroyer⁹, Chen Liu¹¹, Richard Kolesnick⁷, Marcel R. M. van den Brink^{1,3§} & Alan M. Hanash^{1§}

Epithelial regeneration is critical for barrier maintenance and organ function after intestinal injury. The intestinal stem cell (ISC) niche provides Wnt, Notch and epidermal growth factor (EGF) signals supporting Lgr5⁺ crypt base columnar ISCs for normal epithelial maintenance^{1,2}. However, little is known about the regulation of the ISC compartment after tissue damage. Using *ex vivo* organoid cultures, here we show that innate lymphoid cells (ILCs), potent producers of interleukin-22 (IL-22) after intestinal injury^{3,4}, increase the growth of mouse small intestine organoids in an IL-22-dependent fashion. Recombinant IL-22 directly targeted ISCs, augmenting the growth of both mouse and human intestinal organoids, increasing proliferation and promoting ISC expansion. IL-22 induced STAT3 phosphorylation in Lgr5⁺ ISCs, and STAT3 was crucial for both organoid formation and IL-22-mediated regeneration. Treatment with IL-22 *in vivo* after mouse allogeneic bone marrow transplantation enhanced the recovery of ISCs, increased epithelial regeneration and reduced intestinal pathology and mortality from graft-versus-host disease. ATOH1-deficient organoid culture demonstrated that IL-22 induced epithelial regeneration independently of the Paneth cell niche. Our findings reveal a fundamental mechanism by which the immune system is able to support the intestinal epithelium, activating ISCs to promote regeneration.

The epithelial layer in the gastrointestinal tract represents a fundamental line of defence against potential enteric pathogens. Paneth cells contribute to this defence by producing antimicrobial molecules and by providing an epithelial niche for Lgr5⁺ ISCs that maintain the epithelium². ISCs are critical for damage-induced intestinal regeneration⁵, but the mechanisms regulating ISC function and inducing epithelial regeneration after tissue damage remain poorly understood. Furthermore, although epithelial barrier function is a core component of intestinal immunity, little is known about the role of the immune system in regulating the ISC compartment. Group 3 ILCs (ILC3s) are crucial for maintaining gastrointestinal epithelial integrity and barrier function in several experimental models of intestinal injury³. Tissue-resident ILC3s are potent producers of IL-22 after damage, and IL-22 expression is associated with reduced injury in colitis as well as several non-intestinal tissue damage models^{3,4,6–9}. However, although the IL-22 receptor (IL-22R) is present in many epithelial tissues, the specific cellular targets and mechanisms of IL-22 inducing tissue recovery are largely unknown. Using an organoid model of

ex vivo epithelial regeneration¹⁰, we examined whether ILCs and IL-22 could regulate the ISC compartment.

We first sorted mouse small intestine (SI) lamina propria lymphocytes (LPLs), which include both innate and adaptive lymphoid cells capable of producing IL-22 (ref. 4), and cultured them with freshly isolated mouse SI crypts in standard organoid media containing EGF, Noggin and R-spondin-1 (ENR). An IL-23-based cytokine cocktail was included for IL-22 induction. Two-dimensional perimeter tracing (Extended Data Fig. 1a) indicated that co-culture with wild-type LPLs significantly increased organoid size (Fig. 1a). By contrast, LPLs isolated from IL-22-deficient (*Il22*^{−/−}) mice failed to augment organoid size (Fig. 1a). To evaluate the role of ILC3s in organoid growth, SI lamina propria CD45⁺CD3[−]RORγt⁺ ILC3s were isolated from *Rorc*(γt)-GFP (green fluorescent protein) reporter mice and cultured with SI crypts. ILC3s significantly increased SI organoid size, and this was inhibited by an anti-IL-22 neutralizing antibody (Fig. 1b).

Given that IL-22 was essential for ILC-mediated augmentation of organoid size, we focused on studies with recombinant mouse (rm)IL-22. SI crypts cultured with rmIL-22 yielded substantially larger organoids in a concentration-dependent fashion (Fig. 1c, d and Extended Data Fig. 1b). While high concentrations of IL-22 reduced the efficiency of organoid generation from SI crypts, culture with 1–5 ng ml^{−1} rmIL-22 increased organoid size without affecting organoid formation (Extended Data Fig. 1c). IL-22 also increased large intestine organoid size without affecting efficiency (Fig. 1e and Extended Data Fig. 1d), and culture with IL-22 augmented crypt budding in both small and large intestine organoids (Fig. 1f). Furthermore, recombinant human (rh)IL-22 significantly increased the size of human intestinal organoids generated from primary duodenal tissue (Fig. 1g and Extended Data Fig. 1e).

Wnt/β-catenin signalling is essential for ISC maintenance and organoid function *ex vivo*¹⁰. However, we found no evidence of enhanced production of molecules in the Wnt/β-catenin pathway within SI organoids cultured with IL-22, including no difference in expression of WNT3, β-catenin or the downstream target AXIN2 (Extended Data Fig. 1f). Consistent with this, IL-22 could not replace R-spondin-1, an agonist of Wnt/β-catenin signalling, as its removal eliminated SI organoid growth even in the presence of IL-22 (Extended Data Fig. 1g). Additionally, we found no IL-22-induced activation of gene expression in the Notch pathway, which is also critical for ISC maintenance, or activation of gene expression

¹Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ²Department of Pediatrics, University Medical Center Utrecht, 3508 AB Utrecht, The Netherlands. ³Department of Immunology, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ⁴Department of Anatomy and Developmental Biology, Monash University, Clayton 3800, Australia. ⁵Department of Medicine, Weill Cornell Medicine, New York, New York 10021, USA. ⁶Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ⁷Department of Molecular Pharmacology, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ⁸Department of Cancer Biology & Genetics, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. ⁹Department of Medicine, Baylor College of Medicine, Houston, Texas 77030, USA. ¹⁰Department of Hematology, Erasmus University Medical Center, 3000 CA Rotterdam, The Netherlands. ¹¹Department of Pathology, Immunology and Laboratory Medicine, University of Florida College of Medicine, Gainesville, Florida 32610, USA.

*These authors contributed equally to this work.

§These authors jointly supervised this work.

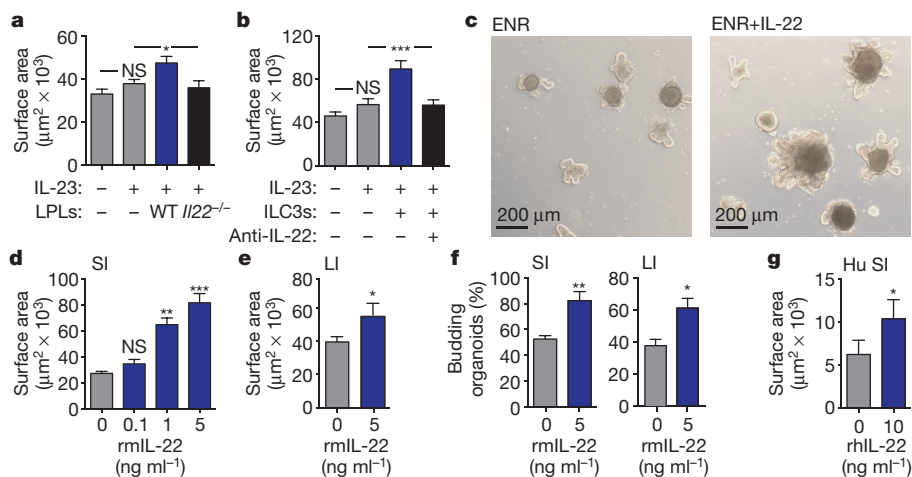


Figure 1 | IL-22 increases growth of intestinal organoids. **a**, Size of SI organoids cultured in ENR with IL-23-containing cytokine cocktail with/without LPLs; $n = 62$ (control), $n = 72$ (IL-23), $n = 29$ (wild-type (WT) LPLs), $n = 34$ (*Il22*^{-/-} LPLs) organoids per group; one of two experiments. **b**, Size of SI organoids cultured with/without ILC3s and anti-IL-22 neutralizing antibody; $n = 47$ (control), $n = 55$ (IL-23), $n = 43$ (ILC3s), $n = 38$ (anti-IL-22) organoids per group; one of two experiments. **c**, SI organoids cultured with/without rmIL-22 (5 ng ml^{-1}) for 7 days. **d**, **e**, Size of organoids cultured with/without rmIL-22 for 7 days; $n = 114$ (control), $n = 50$ (0.1 ng ml^{-1}), $n = 47$ (1 ng ml^{-1}), $n = 44$ (5 ng ml^{-1}) SI

organoids per group (**d**); $n = 115$ (control), $n = 61$ (IL-22) large intestine (LI) organoids per group (**e**). **f**, New crypt formation (budding) of small (day 4) and large (day 7) intestine organoids; $n = 6$ mice per group. **g**, Size of human (Hu) SI organoids cultured with/without rhIL-22 (10 ng ml^{-1}) in standard expansion medium; $n = 38$ (control), $n = 67$ (IL-22) organoids per group. Data are mean and s.e.m.; comparisons performed with *t*-tests (two groups) or analysis of variance (ANOVA) (multiple groups). NS, not significant; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Data combined from at least three independent experiments unless otherwise stated.

for SLIT2 or ROBO1, although they can regulate ISC recovery from damage induced by chemotherapy and radiation¹¹ (Extended Data Fig. 1h). Consistently, there was also no increase in Wnt or Notch pathway gene expression in large intestine organoids (Extended Data Fig. 1i). However, culture with IL-22 increased SI organoid mRNA levels of the innate antimicrobial molecules *Reg3b* and *Reg3g* (Extended Data Fig. 1j), the expression of which is dependent on STAT3 signalling¹².

Little is known about JAK/STAT signalling within ISCs, although it has been reported that STAT3 may be important for ISC maintenance¹³. We evaluated STAT3 signalling in SI organoids and found that IL-22 increased the phosphorylation of STAT3 Tyr705 (Extended Data Fig. 2a). Furthermore, treatment with the STAT3 inhibitor Stattic significantly impaired SI organoid growth (Fig. 2a and Extended Data Fig. 2b). However, IL-22 can also promote epithelial STAT1 signalling¹⁴, which Stattic has inhibitory activity against¹⁵. Indeed, organoid STAT3 and STAT1 were both phosphorylated in response to IL-22, and inhibited by Stattic (Fig. 2b). To determine their relative importance for IL-22-induced epithelial regeneration, we assessed the growth of organoids with genetic deletion of either *Stat1* or *Stat3*. Despite the induction of phosphorylated (p) STAT1 by IL-22, SI crypts from *Stat1*^{-/-} mice demonstrated intact organoid growth and response to IL-22 (Fig. 2c and Extended Data Fig. 2c). As *Stat3*^{-/-} mice are not viable, we next cultured SI crypt cells from *Stat3*^{fl/fl} mice with adenoviral-Cre (adeno-Cre) to delete STAT3. Crypt cells from wild-type mice demonstrated intact organoid growth and IL-22 response despite *in vitro* infection with adeno-Cre (Fig. 2d). Additionally, uninfected *Stat3*^{fl/fl} crypt cells demonstrated normal organoid growth and response to IL-22 (Extended Data Fig. 2d, e). However, crypt cells from *Stat3*^{fl/fl} mice failed to generate organoids after infection with adeno-Cre, and IL-22 failed to recover organoid growth or augment organoid size (Fig. 2d).

Lgr5⁺ ISCs can generate all cell types of mature intestinal epithelium *ex vivo* and *in vivo*^{1,10}. To evaluate whether STAT3 was important for ISCs during tissue damage *in vivo*, we performed a gene set enrichment analysis (GSEA), assessing expression of a published Lgr5⁺ ISC gene signature (Gene Expression Omnibus (GEO) data set GSE33948)¹⁶ in another data set of wild-type (*Stat3*^{fl/fl}) versus epithelial STAT3-deficient (*Stat3*^{fl/fl}; *Villin-Cre*) mice with dextran sulfate sodium (DSS) colitis (GSE15955)¹². Expression of the Lgr5⁺ ISC gene

signature was significantly reduced in STAT3-deficient mice with colitis (Fig. 2e). This was validated with a second independently established Lgr5⁺ ISC gene signature (GEO data set GSE23672)¹⁶, while no significant changes were seen with a negative-control Paneth cell gene signature (GEO data set GSE39915)¹⁷ (Extended Data Fig. 2f, g). Given the induction of pSTAT3 by IL-22 and the importance of STAT3 for ISC gene signature maintenance, we examined the effect of IL-22 during regeneration using purified ISCs. We isolated Lgr5-GFP SI ISCs by fluorescence-activated cell sorting (FACS) and cultured purified ISCs under standard conditions¹⁰ with or without IL-22. IL-22 significantly increased the budding of early organoids after just 4 days (Fig. 2f). Furthermore, as with crypt-derived organoids, 1 ng ml^{-1} rmIL-22 augmented the size of organoids generated from purified ISCs without affecting the efficiency of organoid formation (Fig. 2g and Extended Data Fig. 3a, b).

Consistent with increased size, IL-22 enhanced 5-ethynyl-2'-deoxyuridine (EdU) incorporation in SI organoids, demonstrating increased proliferation (Extended Data Fig. 4a, b). Hoechst staining revealed an IL-22-dependent increase in G2/M populations after 2 days in culture (Fig. 2h), and IL-22 treatment rapidly reduced expression of key cell cycle checkpoint molecules *Cdkn1a* and *Cdkn2d* in both small and large intestine organoids (Extended Data Fig. 4c, d). Furthermore, IL-22 expanded Lgr5-GFP^{high} ISCs in SI organoids (Fig. 2i) and increased expansion of SI organoids over several passages in culture (Fig. 2j). Next, we evaluated the ISC compartment after radiation injury. Pre-treatment with rmIL-22 increased the percentage of dissociated SI crypt cells that were viable in culture after *ex vivo* irradiation, as measured by MTT reduction (Extended Data Fig. 5a). Consistent with this, IL-22 treatment increased the number of organoids that could grow from single cells 2 days after irradiation (Extended Data Fig. 5b). This was more evident with increasing doses of irradiation, and protection was present even 7 days after irradiation (Extended Data Fig. 5c). Accordingly, irradiation was found to increase the expression of *Il22ra1* within intestinal crypts (Extended Data Fig. 5d).

We next evaluated the effect of IL-22 *in vivo* after tissue damage using a clinically relevant mouse graft-versus-host disease (GVHD) model. T-cell-depleted (TCD) marrow from LP mice was transplanted with or without purified LP T cells into lethally irradiated C57BL/6 (B6) recipients (H-2^b into H-2^b). Mice receiving allogeneic T cells for

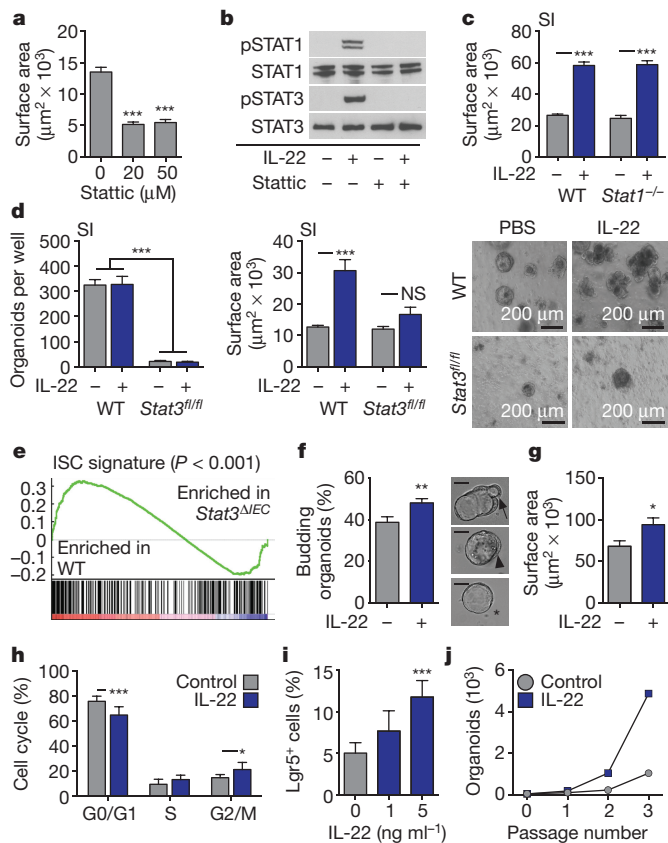


Figure 2 | IL-22 activates organoid STAT3 signalling and augments ISC regeneration. **a**, SI organoid size, after 4 days with/without Statistic; $n = 174$ (control), $n = 134$ (20 μM), $n = 102$ (50 μM) organoids per group. **b**, Crypt pSTAT1 western blots after 30 min incubation with rmIL-22 (5 ng ml^{-1}) with/without Statistic; one of three experiments. **c**, Size of day 7 wild-type and $\text{Stat1}^{-/-}$ SI organoids with/without rmIL-22 (5 ng ml^{-1}); $n = 821$ (wild type), $n = 503$ (wild type plus IL-22), $n = 432$ ($\text{Stat1}^{-/-}$), $n = 269$ ($\text{Stat1}^{-/-}$ plus IL-22) organoids per group. **d**, Day 5 $\text{Stat3}^{\text{fl/fl}}$ SI organoids cultured with adeno-Cre with/without rmIL-22 (5 ng ml^{-1}); numbers per well, $n = 6$ wells per group; size, $n = 253$ (wild type), $n = 49$ (wild type plus IL-22), $n = 38$ ($\text{Stat3}^{\text{fl/fl}}$), $n = 38$ ($\text{Stat3}^{\text{fl/fl}}$ plus IL-22) organoids per group; images (right) representative of three experiments. **e**, GSEA of ISC signature genes in wild-type versus $\text{Stat3}^{\text{fl/fl}}$, Villin-Cre ($\text{Stat3}^{\Delta\text{IEC}}$) mice with DSS colitis; one analysis, nominal P value shown. **f**, **g**, Organoids from sorted SI Lgr5-GFP⁺ ISCs cultured with/without rmIL-22 (1 ng ml^{-1}). **f**, Organoid budding, percentage of total organoids per well (day 4, $n = 11$ wells per group). Representative images (right) of early budding indicate: early organoid without budding (asterisk); polarization before budding (arrowhead); budding at site of polarization (arrow). Scale bars, 50 μm . **g**, Organoid area (day 13), $n = 54$ organoids per group. **h**, Cell cycle FACS of SI organoid cells cultured with/without rmIL-22 (5 ng ml^{-1}); $n = 7$ mice per group. **i**, FACS analysis of Lgr5-GFP^{high} ISCs in organoids cultured with/without rmIL-22; $n = 6$ mice per group. **j**, Organoid expansion with serial passaging with/without rmIL-22 (1 ng ml^{-1}); one of two experiments. Data are mean and s.e.m.; comparisons performed with t -tests (two groups) or ANOVA (multiple groups); * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Data combined from at least two independent experiments unless otherwise stated. For western blot source data, see Supplementary Fig. 1.

GVHD induction were treated with 4 μg rmIL-22 or PBS daily via intraperitoneal (i.p.) injection starting 7 days after bone marrow transplantation (BMT). Treatment with IL-22 reduced histopathological evidence of GVHD in the small and large intestine (Fig. 3a), including a reduction in apoptosis within crypt epithelium (Extended Data Fig. 6a, b). GVHD pathology was reduced despite an intact alloimmune response, as evidenced by similar T cell subset distribution, activation markers and gut homing molecules, as well as similar systemic and gastrointestinal expression of inflammatory cytokines (Extended Data Fig. 6c, d). However, IL-22 treatment did increase SI expression

of *Reg3b* and *Reg3g* mRNA (Fig. 3b). Consistent with previous findings¹⁸, we found that REG3 β was primarily expressed by enterocytes in allogeneic BMT recipients, including after treatment with IL-22 (Extended Data Fig. 6e).

GVHD is associated with a loss of both ISCs^{19,20} and niche-forming Paneth cells^{21,22}, and T-cell-replete BMT led to a significant loss of Lgr5-LacZ⁺ SI ISCs 3 weeks after transplantation (Fig. 3c). However, IL-22 treatment increased the recovery of Lgr5⁺ ISCs (Fig. 3c). This was associated with increased regeneration as evidenced by increased crypt height, including the transit-amplifying compartment (Fig. 3d). Paneth cells support Lgr5⁺ ISCs through the delivery of Wnt, Notch ligand and EGF signals². Additionally, IL-22 is thought to regulate Paneth cell production of innate antimicrobial molecules. We thus proposed that IL-22 could support ISC recovery after BMT by improving the function of the stem-cell niche. Consistent with previous clinical and experimental studies^{21–23}, minor antigen-mismatched BMT led to a reduction in Paneth cells 3 weeks after transplantation (Fig. 3e). However, IL-22 administration did not increase recovery of Paneth cells (Fig. 3e), mRNA expression of *Wnt3* or *Egf* (Extended Data Fig. 6f, g), or expression of Notch ligand target *Hes1* (Extended Data Fig. 6h). Furthermore, although stroma can support ISC Wnt signalling *in vivo* independently of Paneth cells²⁴, we found no change in expression of stromal R-spondin-3 after IL-22 treatment post-BMT, and no change in expression of Wnt pathway genes regardless of the upstream source (Extended Data Fig. 6i–k).

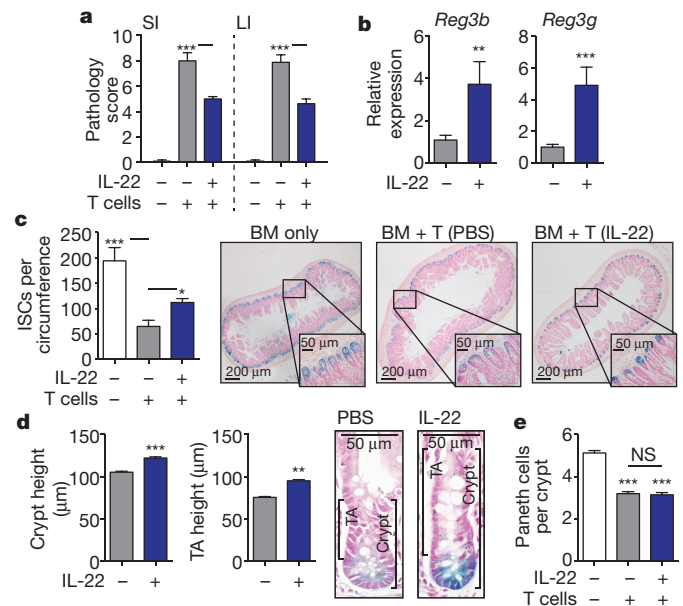


Figure 3 | IL-22 reduces intestinal pathology and increases ISC recovery after *in vivo* tissue damage. LP into B6 BMT; recipients treated daily with PBS or 4 μg rmIL-22 i.p. starting 7 days after BMT. **a**, Intestinal GVHD histopathology score, 3 weeks after BMT; $n = 10$ (TCD bone marrow (BM) only), $n = 9$ (BM plus T (PBS)), $n = 8$ (BM plus T (IL-22)) mice per group; Kruskal–Wallis analysis. **b**, qPCR of *Reg3b* and *Reg3g* in SI tissue 3 weeks after BMT; $n = 9$ (PBS) and $n = 10$ (IL-22) mice per group; Mann–Whitney U analysis. **c**, **d**, B6 Lgr5-LacZ recipients. **c**, SI ISC frequency 3 weeks after BMT; Kruskal–Wallis analysis of $n = 8$ (TCD BM only), $n = 20$ (BM plus T (PBS)), or $n = 20$ (BM plus T (IL-22)) independent sections (four sections per recipient from 2–5 mice per group); one of two experiments. **d**, Crypt and transit-amplifying (TA) heights 3 weeks after BMT, representative images on right; t -test analyses of $n = 285$ (PBS) versus $n = 324$ (IL-22) crypts, and $n = 168$ (PBS) versus $n = 224$ (IL-22) transit-amplifying compartments (one section per mouse, >10 mice per group). **e**, SI lysozyme⁺ Paneth cell frequency; Kruskal–Wallis analysis of $n = 73$ (TCD BM only), $n = 89$ (BM plus T (PBS)), and $n = 88$ (BM plus T (IL-22)) crypts (5–8 mice per group). Data are mean and s.e.m.; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Data combined from at least two independent experiments unless otherwise stated.

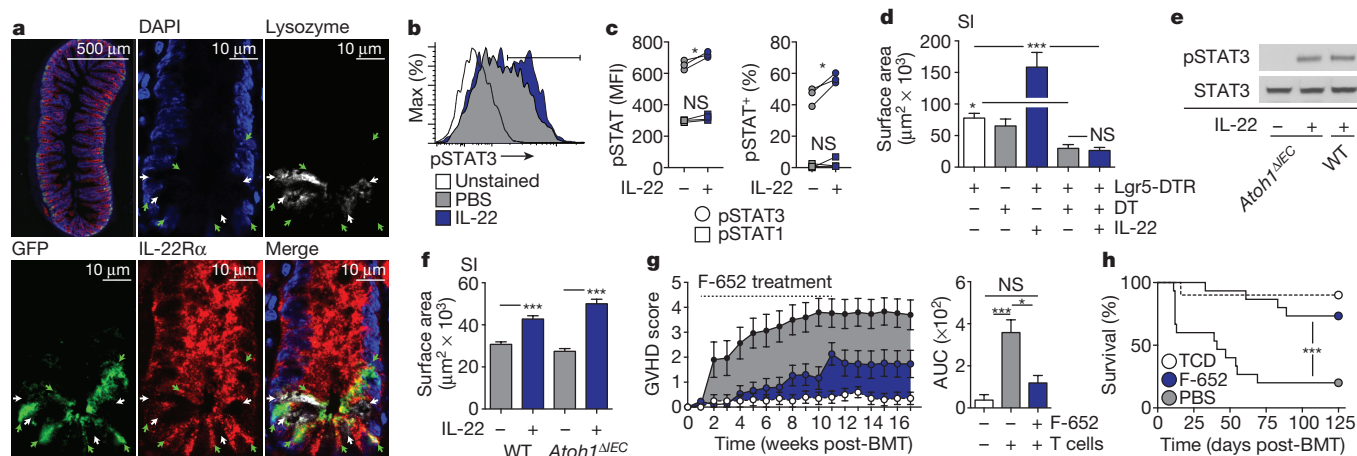


Figure 4 | IL-22 directly promotes ISC-dependent epithelial regeneration. **a**, Immunofluorescent staining of IL-22Rα1, GFP and lysozyme in SI sections from *Lgr5-GFP* mice; green arrows, *Lgr5-GFP*⁺ ISCs; white arrows; lysozyme⁺ Paneth cells. DAPI, 4',6-diamidino-2-phenylindole. **b**, **c**, Phosflow analysis of *Lgr5-GFP*⁺ SI crypt cells after 30 min with/without rmIL-22 (20 ng ml⁻¹). **b**, pSTAT3 histogram; representative of four experiments. **c**, pSTAT median fluorescence intensity (MFI) and percentage pSTAT⁺; *n* = 3 mice per group; representative of two experiments. **d**, Size of wild-type and *Lgr5-DTR* day 5 SI organoids cultured with diphtheria toxin (DT; 1 ng μl⁻¹) to deplete *Lgr5*⁺ cells with/without rmIL-22 (5 ng ml⁻¹); one of three experiments; *n* = 65 (wild type), *n* = 25 (wild type plus DT), *n* = 28 (DTR plus IL-22), *n* = 18 (DTR plus diphtheria toxin), *n* = 40 (DTR, diphtheria toxin and IL-22) organoids per group. **e**, **f**, Paneth-cell-deficient *Atoh1*^{ΔIEC} SI

Given that IL-22 treatment appeared to improve ISC numbers without improving niche function in BMT recipients, we sought to evaluate how IL-22 was targeting the ISC compartment. Consistent with the *in vivo* findings, IL-22 had no effect on Paneth cell frequency or α-defensin-1 expression within SI organoids cultured *ex vivo* (Extended Data Fig. 7a, b). Immunofluorescent staining for IL-22Rα1 and the Paneth cell marker lysozyme using SI sections from *Lgr5-GFP* reporter mice indicated substantial IL-22R staining within crypts and enterocytes at the villous base, but not on lysozyme⁺ Paneth cells (Fig. 4a). Using flow cytometry, there was also little evidence for expression of IL22Rα1 on Paneth cells at baseline or after radiation injury and no evidence of pSTAT3 in Paneth cells in response to IL-22 (Extended Data Fig. 7c–e). By contrast, IL22Rα1 was identified in the transit-amplifying progenitor compartment and on *Lgr5-GFP*⁺ ISCs (Fig. 4a). IL-22R expression in *Lgr5*⁺ cells was confirmed by quantitative PCR (qPCR) after sorting for *Lgr5-GFP*⁺ cells (Extended Data Fig. 8a, b). Furthermore, SI crypt cells from *Lgr5-GFP* reporter mice demonstrated increased STAT3 Tyr705 phosphorylation within GFP⁺ cells after incubation with IL-22, indicating functional IL-22R signalling in ISCs (Fig. 4b, c). STAT3 phosphorylation was a specific response, as there was no effect of IL-22 on pSTAT1 in *Lgr5*⁺ cells (Fig. 4c).

IL-22R expression and STAT3 phosphorylation suggested that IL-22 might promote regeneration via direct targeting of ISCs. To investigate the role of ISCs and Paneth cells in IL-22-mediated regeneration further, we assessed organoid growth when either cell population was depleted. Treatment of transgenic mice expressing the diphtheria toxin receptor (*Lgr5-DTR*) with diphtheria toxin leads to a rapid deletion of *Lgr5*⁺ cells²⁵. Deletion of *Lgr5*⁺ cells *ex vivo* by culturing *Lgr5-DTR* SI organoids with diphtheria toxin impaired epithelial regeneration as evidenced by a reduction in organoid size and efficiency (Fig. 4d and Extended Data Fig. 9a). Although IL-22 increased the size of *Lgr5-DTR* organoids cultured without diphtheria toxin, IL-22 failed to increase the size or maintain the numbers of *Lgr5-DTR* organoids cultured with diphtheria toxin (Fig. 4d and Extended Data Fig. 9a), indicating that *Lgr5*⁺ cells are essential for IL-22-mediated epithelial regeneration.

organoids cultured in WNT3-supplemented ENR with/without rmIL-22 (5 ng ml⁻¹). **e**, STAT3 western blots after 30 min culture with rmIL-22; one of four experiments. **f**, Day 7 organoid size; *n* = 466 (wild type), *n* = 531 (wild type plus IL-22), *n* = 197 (*Atoh1*^{ΔIEC}), *n* = 491 (*Atoh1*^{ΔIEC} plus IL-22) organoids per group. **g**, **h**, LP into B6 BMT, with/without F-652 (100 μg kg⁻¹ subcutaneous, every other day starting day 7 after BMT, 10-week course); *n* = 10 (TCD BM only), *n* = 15 (BM plus T (PBS)), *n* = 15 (BM plus T (IL-22)). **g**, Clinical signs of GVHD and area under the curve (AUC) analysis of GVHD scoring. **h**, Percentage survival. Data are mean and s.e.m.; comparisons performed with *t*-tests (two groups), ANOVA (multiple groups), or log-rank analysis (**h**); **P* < 0.05, ****P* < 0.001. Data combined from at least two independent experiments unless otherwise stated. For western blot source data, see Supplementary Fig. 1.

We next investigated the functional importance of Paneth cells for IL-22-mediated regeneration by culturing organoids with IL-22 after inducible Paneth cell depletion. Paneth cells were deleted *in vivo* after tamoxifen treatment of *Atoh1*^{fl/fl}; *Villin-Cre*^{ERT2} (*Atoh1*^{ΔIEC}) mice²⁶. IL-22 led to robust STAT3 phosphorylation within *Atoh1*^{ΔIEC} SI organoids, confirming that Paneth cells were not essential for IL-22-mediated intracellular signalling, and *Atoh1*^{ΔIEC} organoids demonstrated an intact growth response to IL-22 (Fig. 4e, f and Extended Data Fig. 9b). Additionally, we found that IL-22 could augment the size of organoids cultured without EGF (Extended Data Fig. 9c, d). These findings indicated that the Paneth cell niche was not required for IL-22-mediated epithelial regeneration, and IL-22 could promote the growth of organoids cultured without addition of the niche-derived growth factor EGF.

Recent reports have suggested that T-cell-derived IL-22 may contribute to GVHD, as might peri-transplant administration of IL-22 to MHC-mismatched BMT recipients^{27,28}. However, IL-22-producing ILCs are eliminated in GVHD¹⁹, ILC deficiency is associated with increased clinical GVHD²⁹, and gastrointestinal damage may be central to the pathogenesis of systemic GVHD³⁰. We thus proposed that stimulating regeneration with IL-22 after the initiation of GVHD-related tissue damage may be therapeutically beneficial. Given the improved pharmacological stability of Fc-fusion molecules, we evaluated the potential of F-652, a rhIL-22-dimer and Fc-fusion protein, for treatment of systemic GVHD. First, we found that F-652 had activity in mouse epithelial regeneration, augmenting the growth of both small and large intestine organoids without evidence of toxicity (Extended Data Fig. 10a–d). Second, treatment of *Lgr5-LacZ* reporter mice with F-652 significantly protected SI *Lgr5*⁺ crypt cells from radiation injury *in vivo* (Extended Data Fig. 10e, f). We next investigated an early intervention model for GVHD, treating allogeneic BMT recipients (LP into B6) with F-652 starting 1 week after transplantation. Mice treated with F-652 demonstrated reduced systemic signs of GVHD and GVHD-related mortality compared to PBS-treated controls (Fig. 4g, h).

In summary, we found that IL-22 links immunity to epithelial regeneration by acting directly on ISCs. Purified ILCs enhanced organoid

growth in an IL-22-dependent fashion, and IL-22 augmented ISC-mediated epithelial regeneration, promoting cell cycle progression, epithelial proliferation and regeneration of the ISC pool. IL-22 induced STAT3 phosphorylation in Lgr5⁺ ISCs, and while IL-22 may not be its sole regulator in ISCs, STAT3 was essential for organoid growth and IL-22-dependent epithelial regeneration. Paneth cells, in contrast, were not required for IL-22-driven regeneration. Given the activation of IL-22 production and the upregulation of crypt IL-22R expression after tissue damage, these findings indicate that IL-22 contributes to damage-induced regulation of the ISC compartment. We conclude that in addition to the stromal and epithelial components of the ISC niche that are essential for normal epithelial maintenance, IL-22 provides evidence for an immunological contribution to the ISC niche that is activated to restore the epithelium after tissue injury. By acting directly on epithelial stem cells, the immune system is thus able to regulate intestinal regeneration and support the fundamental defence system provided by the integrity of the epithelial barrier.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 October 2014; accepted 18 November 2015.

Published online 9 December 2015.

- Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449**, 1003–1007 (2007).
- Sato, T. *et al.* Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* **469**, 415–418 (2011).
- Sonnenberg, G. F. & Artis, D. Innate lymphoid cells in the initiation, regulation and resolution of inflammation. *Nature Med.* **21**, 698–708 (2015).
- Dudakov, J. A., Hanash, A. M. & van den Brink, M. R. Interleukin-22: immunobiology and pathology. *Annu. Rev. Immunol.* **33**, 747–785 (2015).
- Metcalfe, C., Kljavin, N. M., Ybarra, R. & de Sauvage, F. J. Lgr5⁺ stem cells are indispensable for radiation-induced intestinal regeneration. *Cell Stem Cell* **14**, 149–159 (2014).
- Zheng, Y. *et al.* Interleukin-22 mediates early host defense against attaching and effacing bacterial pathogens. *Nature Med.* **14**, 282–289 (2008).
- Zenewicz, L. A. *et al.* Interleukin-22 but not interleukin-17 provides protection to hepatocytes during acute liver inflammation. *Immunity* **27**, 647–659 (2007).
- Aujla, S. J. *et al.* IL-22 mediates mucosal host defense against Gram-negative bacterial pneumonia. *Nature Med.* **14**, 275–281 (2008).
- Dudakov, J. A. *et al.* Interleukin-22 drives endogenous thymic regeneration in mice. *Science* **336**, 91–95 (2012).
- Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures *in vitro* without a mesenchymal niche. *Nature* **459**, 262–265 (2009).
- Zhou, W. J., Geng, Z. H., Spence, J. R. & Geng, J. G. Induction of intestinal stem cells by R-spondin 1 and Slit2 augments chemoradioprotection. *Nature* **501**, 107–111 (2013).
- Pickert, G. *et al.* STAT3 links IL-22 signaling in intestinal epithelial cells to mucosal wound healing. *J. Exp. Med.* **206**, 1465–1472 (2009).
- Matthews, J. R., Sansom, O. J. & Clarke, A. R. Absolute requirement for STAT3 function in small-intestine crypt stem cell survival. *Cell Death Differ.* **18**, 1934–1943 (2011).
- Hernández, P. P. *et al.* Interferon- λ and interleukin 22 act synergistically for the induction of interferon-stimulated genes and control of rotavirus infection. *Nature Immunol.* **16**, 698–707 (2015).
- Schust, J., Sperl, B., Hollis, A., Mayer, T. U. & Berg, T. Stattic: a small-molecule inhibitor of STAT3 activation and dimerization. *Chem. Biol.* **13**, 1235–1242 (2006).
- Muñoz, J. *et al.* The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent ‘+4’ cell markers. *EMBO J.* **31**, 3079–3091 (2012).
- van Es, J. H. *et al.* Dll1⁺ secretory progenitor cells revert to stem cells upon crypt damage. *Nature Cell Biol.* **14**, 1099–1104 (2012).
- Eriguchi, Y. *et al.* Reciprocal expression of enteric antimicrobial proteins in intestinal graft-versus-host disease. *Biol. Blood Marrow Transplant.* **19**, 1525–1529 (2013).
- Hanash, A. M. *et al.* Interleukin-22 protects intestinal stem cells from immune-mediated tissue damage and regulates sensitivity to graft versus host disease. *Immunity* **37**, 339–350 (2012).
- Takashima, S. *et al.* The Wnt agonist R-spondin1 regulates systemic graft-versus-host disease by protecting intestinal stem cells. *J. Exp. Med.* **208**, 285–294 (2011).
- Jenq, R. R. *et al.* Regulation of intestinal inflammation by microbiota following allogeneic bone marrow transplantation. *J. Exp. Med.* **209**, 903–911 (2012).
- Eriguchi, Y. *et al.* Graft-versus-host disease disrupts intestinal microbial ecology by inhibiting Paneth cell production of α -defensins. *Blood* **120**, 223–231 (2012).

- Levine, J. E. *et al.* Low Paneth cell numbers at onset of gastrointestinal graft-versus-host disease identify patients at high risk for nonrelapse mortality. *Blood* **122**, 1505–1509 (2013).
- Kabiri, Z. *et al.* Stroma provides an intestinal stem cell niche in the absence of epithelial Wnts. *Development* **141**, 2206–2215 (2014).
- Tian, H. *et al.* A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable. *Nature* **478**, 255–259 (2011).
- Durand, A. *et al.* Functional intestinal stem cells after Paneth cell ablation induced by the loss of transcription factor Math1 (Atoh1). *Proc. Natl Acad. Sci. USA* **109**, 8965–8970 (2012).
- Couturier, M. *et al.* IL-22 deficiency in donor T cells attenuates murine acute graft-versus-host disease mortality while sparing the graft-versus-leukemia effect. *Leukemia* **27**, 1527–1537 (2013).
- Zhao, K. *et al.* Interleukin-22 aggravates murine acute graft-versus-host disease by expanding effector T cell and reducing regulatory T cell. *J. Interferon Cytokine Res.* **34**, 707–715 (2014).
- Munneke, J. M. *et al.* Activated innate lymphoid cells are associated with a reduced susceptibility to graft-versus-host disease. *Blood* **124**, 812–821 (2014).
- Hill, G. R. & Ferrara, J. L. The primacy of the gastrointestinal tract as a target organ of acute graft-versus-host disease: rationale for the use of cytokine shields in allogeneic bone marrow transplantation. *Blood* **95**, 2754–2759 (2000).

Supplementary Information is available in the online version of the paper.

Acknowledgements We gratefully acknowledge the technical assistance of the MSKCC Research Animal Resource Center and Molecular Cytology Core Facility. We also thank H. Clevers, H. Farin, S. Middendorp, C. Wiegierinck, J. van Es, M. van de Wetering, N. Sasaki, J. Sun and M. Li for their advice and critical evaluation of our work. This research was supported by National Institutes of Health award numbers K08-HL115355 (A.M.H.), R01-HL125571 (A.M.H.), R01-HL069929 (M.R.M.vdB.), R01-AI100288 (M.R.M.vdB.), R01-AI080455 (M.R.M.vdB.), R01-AI101406 (M.R.M.vdB.), P01-CA023766/Project 4 (R. J. O'Reilly/M.R.M.vdB.), K99-CA176376 (J.A.D.) and P30-CA008748 (MSKCC Core Grant). Support was also received from the US National Institute of Allergy and Infectious Diseases (NIAID contract HHSN272200900059C), the European Union (award GC220918, C. Blackburn), The Experimental Therapeutics Center of MSKCC funded by Mr William H. Goodwin and Mrs Alice Goodwin, The Lymphoma Foundation, Alex's Lemonade Stand, The Geoffrey Beene Cancer Research Center at MSKCC, The Susan and Peter Solomon Divisional Genomics Program, MSKCC Cycle for Survival, and The Lucille Castori Center for Microbes, Inflammation & Cancer. T.C. was supported by Innovational Research Incentives Scheme Vidi grant 91710377 from the Netherlands Organization for Scientific Research (Zon-MW), and M.R.-H. was supported by the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007–2013 under REA grant agreement no. 289720. A.M.M. was supported by the Bio Medical Exchange Program of the Deutscher Akademischer Austauschdienst. C.A.L. was supported by Dutch Cancer Society clinical fellowship grant 2013-5883 and by a mobility grant from the University Medical Center Utrecht. J.A.D. was supported by a C. J. Martin fellowship from the Australian National Health and Medical Research Council, a Scholar Award from the American Society of Hematology, and the Mechtild Harf Research Grant from the DKMS Foundation for Giving Life. A.M.H. was supported by a Scholar Award from the American Society of Hematology, a New Investigator Award from the American Society for Blood and Marrow Transplantation, and the Amy Strelzer Manasevit Research Program. A provisional patent application has been filed on the use of IL-22 and F-652 as ISC growth factors (US 61/901,151) with A.M.H., C.A.L. and M.R.M.vdB. listed as inventors.

Author Contributions C.A.L. and M.C. designed and performed organoid experiments. A.M.M. and M.H.O. performed and analysed *in vivo* experiments. J.A.D., R.R.J. and E.V. provided input and helped with various assays. L.F.Y., O.M.S. and G.L. performed and monitored bone marrow transplants and maintained the mouse colonies. J.A.I. assisted with organoid quantification. Y.-Y.F. analysed crypt sizes and confocal microscopy. S.T. assisted with ILC co-culture experiments. G.H., M.L.M. and R.K. assisted with ISC isolation and *in vivo* ISC quantification experiments and provided reagents and expertise. K.P.O. and L.D. assisted with adeno-Cre experiments and optimizing various assays. Y.-H.L. and N.F.S. assisted with Paneth cell deficiency experiments. M.M. and E.E.N. performed the GSEA analyses and assisted with reagents and resources. M.R.-H. performed PCR analyses on purified stem cells and immune cells under the guidance of T.C., and C.L. analysed intestinal histopathology. M.R.M.vdB. and A.M.H. supervised the research. All authors contributed to experimental design, interpretation and manuscript editing.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.M.H. (hanasha@mskcc.org).

METHODS

Mice. C57BL/6 (B6, H-2^b) and LP (H-2^b) mice were obtained from Jackson Laboratory. B6 Lgr5-LacZ and B6 Lgr5-gfp-ires-CreERT2 (Lgr5-GFP) mice were provided by H. Clevers^{1,10}. Mouse maintenance and procedures were done in accordance with the institutional protocol guideline of the Memorial Sloan Kettering Cancer Center (MSKCC) Institutional Animal Care and Use Committee. Mice were housed in micro-isolator cages, five per cage, in MSKCC pathogen-free facilities, and received standard chow and autoclaved sterile drinking water. To adjust for differences in weight and intestinal flora among other factors, identical mice were purchased from Jackson and then randomly distributed over different cages and groups by a non-biased technician who had no insight or information about the purpose or details of the experiment. The investigations assessing clinical outcome parameters were performed by non-biased technicians with no particular knowledge or information regarding the hypotheses of the experiments and no knowledge of the specifics of the individual groups.

Crypt isolation and cell dissociation. Isolation of intestinal crypts and the dissociation of cells for flow cytometry analysis were largely performed as previously described¹⁰. In brief, after euthanizing the mice with CO₂ and collecting small and large intestines, the organs were opened longitudinally and washed with PBS. To dissociate the crypts, small intestine was incubated at 4°C in EDTA (10 mM) for 15 min and then in EDTA (5 mM) for an additional 15 min. Large intestine was incubated in collagenase type 4 (Worthington) for 30 min at 37°C to isolate the crypts. To isolate single cells from small and large intestine crypts, the pellet was further incubated in 1 × TrypLE express (Gibco, Life Technologies) supplemented with 0.8 kU ml⁻¹ DNaseI (Roche).

Organoid culture. For mouse organoids, depending on the experiments, 200–400 crypts per well were suspended in Matrigel composed of 25% advanced DMEM/F12 medium (Gibco) and 75% growth-factor-reduced Matrigel (Corning). After the Matrigel polymerized, complete ENR medium containing advanced DMEM/F12 (Sigma), 2 mM Glutamax (Invitrogen), 10 mM HEPES (Sigma), 100 U ml⁻¹ penicillin, 100 µg ml⁻¹ streptomycin (Sigma), 1 mM N-acetyl cysteine (Sigma), B27 supplement (Invitrogen), N2 supplement (Invitrogen), 50 ng ml⁻¹ mouse EGF (PeproTech), 100 ng ml⁻¹ mouse Noggin (PeproTech) and 10% human R-spondin-1-conditioned medium from R-spondin-1-transfected HEK 293T cells³¹ was added to small intestine crypt cultures¹⁰. For experiments evaluating organoid budding, the concentration of R-spondin-1 was lowered to 1.25–5%. For mouse large intestine, crypts were cultured in 'WENR' medium containing 50% WNT3a-conditioned medium in addition to the aforementioned proteins and 1% BSA (Sigma), and supplemented with SB202190 (10 µM, Sigma), ALK5 inhibitor A83-01 (500 nM, Tocris Bioscience) and nicotinamide (10 mM, Sigma). Media was replaced every 2–3 days. Along with medium changes, treatment wells received different concentrations of rmIL-22 (Genscript). We also tested the effects of F-652 (Generon Corporation). In some experiments, organoids from crypts were cultured in the presence of Stattic (Tocris Bioscience). For passaging of organoids, after 5–7 days of culture, organoids were passaged by mechanically disrupting with a seropipet and cold media to depolymerize the Matrigel and generate organoid fragments. After washing away the old Matrigel by spinning down at 600 r.p.m., organoid fragments were replated in liquid Matrigel.

ISCs were isolated from Lgr5-GFP mice using a modified crypt isolation protocol with 20 min of 30 mM EDTA^{32,33} followed by several strainer steps and a 5-min incubation with TrypLE and 0.8 kU ml⁻¹ DNaseI under minute-to-minute vortexing to make a single-cell suspension. The Lgr5-GFP^{high} cells were isolated by FACS. Approximately 5,000 ISCs were plated in 30 µl Matrigel and cultured in WENR media containing Rho-kinase/ROCK inhibitor Y-27632 (10 µM, Tocris Bioscience) and Jagged1 (1 µM, Anaspec). Starting from day 4, ISC were cultured without Wnt.

For lymphocyte co-culture experiments, ILCs were isolated from the small intestine lamina propria. Washed small intestine fragments were incubated in EDTA/IEL solution (1 × PBS with 5% FBS, 10 mM HEPES buffer, 1% penicillin/streptomycin (Corning), 1% L-glutamine (Gibco), 1 mM EDTA and 1 mM dithiothreitol (DTT)) in a 37°C shaker for 15 min. The samples were strained (100 µm) and put in a Collagenase solution (RPMI 1640, 5% FCS, 10 mM HEPES, 1% penicillin/streptomycin, 1% glutamine, 1 mg ml⁻¹ collagenase D (Roche) and 1 U ml⁻¹ DNaseI (Roche) and incubated twice for 10 min in a 37°C shaker. Afterwards, the samples were centrifuged at 1,500 r.p.m. for 5 min and washed with RPMI solution without enzymes. After several washes, the cell suspension was transferred into a 40% Percoll solution (in PBS), which is overlaid on an 80% Percoll solution. After spinning the interface containing the lamina propria, mononuclear cells were aspirated and washed in medium. The cell suspension was then stained with extracellular markers and Topro3 for viability. Topro3⁺CD45⁺CD11b⁺CD11c⁺CD90⁺ LPLs from B6 wild-type and *Il22*^{-/-} mice and Topro3⁺CD45⁺CD3⁺RORγ⁺ ILC3s³⁴ from Rorc(γ)-GFP⁺ mice (Jackson) were sorted for co-cultures with SI crypts.

(For antibodies used, see Supplementary Table 1.) To activate and maintain LPLs and ILCs in culture, rmIL-2 (1,000 U ml⁻¹), rmIL-15 (10 ng ml⁻¹), rmIL-7 (50 ng ml⁻¹) and rmIL-23 (50 ng ml⁻¹) were added to the ENR medium in co-culture experiments. We have also performed co-cultures with addition of only rmIL-23 (50 ng ml⁻¹) to ENR media. LPLs and SI crypts were cultured in Matrigel with a 7:1 LPL:crypt ratio; ILCs and crypts were cultured in Matrigel with a 25:1 ILC:crypt ratio. Co-cultures were compared to crypts cultured in ENR plus cytokines without LPLs or ILCs present. A neutralizing monoclonal antibody against IL-22 (8E11, Genentech)³⁵ was used to abrogate IL-22-specific effects of ILCs.

For specific experiments, organoids were cultured from fresh crypts obtained from specific genetically modified mice, such as the *Stat1*^{-/-} mice (129S6/SvEv-Stat1 tm1Rds, Taconic) and *Stat3*^{fl/fl} mice (Jackson). Organoids from *Stat3*^{fl/fl} mice that had been grown for 7 days were dissociated as single cells and incubated with adenoviral-Cre (University of Iowa) to cause the deletion of *Stat3* from floxed organoid cells. Frozen passaged organoids from *Lgr5*^{DTR} (Lgr5-DTR)²⁵ mice were used to culture organoids in which Lgr5⁺ stem cells could be depleted with daily administration of diphtheria toxin (1 ng µl⁻¹).

For Paneth-cell-deficient organoid cultures, frozen crypts from *Atoh1*^{ΔIEC} mice³⁶ depleted of Paneth cells were used to culture organoids. As previously described³⁶, *Atoh1*^{ΔIEC} mice (and littermate controls) were given an intraperitoneal injection of tamoxifen (1 mg per mouse, Sigma, dissolved in corn oil) for 5 consecutive days to achieve deletion of ATOH1 from intestinal epithelium. Animals were euthanized on day 7 after the first injection, and intestinal crypts were isolated and frozen in 10% dimethylsulfoxide (DMSO) and 90% FBS.

To investigate the effect of IL-22 on human small intestine, we generated human duodenal organoids from banked frozen organoids (>passage 7) that had been previously generated from biopsies obtained during duodenoscopy of three independent healthy human donors. All human donors had been investigated for coeliac disease, but turned out to have normal pathology. All provided written informed consent to participate in this study according to a protocol reviewed and approved by the review board of the UMC Utrecht, The Netherlands (protocol STEM study, METC 10-402/K). Human organoids were cultured in 10 µl Matrigel drops in expansion medium containing WENR with 10 nM SB202190, 500 nM A83-01 and 10 mM nicotinamide. For IL-22 stimulation experiments, rhIL-22 10 ng ml⁻¹ (Genscript) was added daily. For the purpose of size measurements at day 6, organoids were passaged as single cells.

Where applicable, organoid cultures were performed using conditioned media containing R-spondin-1 and WNT3a produced by stably transfected cell lines. R-spondin-1-transfected HEK293T cells³¹ were provided by C. Kuo. WNT3a-transfected HEK293T cells were provided by H. Clevers (patent WO2010090513A2). Cell lines were tested for mycoplasma and confirmed to be negative.

Organoid measurement. For size evaluation, the surface area of organoid horizontal cross sections was measured. If all organoids in a well could not be measured, several random non-overlapping pictures were acquired from each well using a Zeiss Axio Observer Z1 inverted microscope and then analysed using MetaMorph or ImageJ software. Organoid perimeters for area measurements have been defined manually and by automated determination using the Analyze Particle function of ImageJ software, with investigator verification of the automated determinations, as automated measurements allowed for unbiased analyses of increased numbers of organoids. For automated size measurements, the threshold for organoid identification was set based on monochrome images. The sizes of the largest and smallest organoids in the reference well were measured manually, and their areas were used as the reference values for setting the minimal and maximal particle sizes. Organoids touching the edge of the images were excluded from the counting. After 5–7 days in culture, total organoid numbers per well were counted by light microscopy to evaluate growth efficiency. All organoid numbers were counted manually in this fashion except for the organoid counts presented in Extended Data Fig. 5b, which were counted using automated ImageJ analysis, as these organoids were too numerous to count manually. To compare organoid efficiency in different conditions, combining experiments with different organoid numbers, the percentage of organoids relative to the number of organoids in ENR-control (rmIL-22 0 ng ml⁻¹) was calculated. The efficiency from sorted ISCs was presented as the percentage of cells forming organoids per number of seeded cells.

BMT. BMT procedures were performed as previously described³⁷. A minor histocompatibility antigen-mismatched BMT model (LP into B6; H-2^b into H-2^b) was used. Female B6 wild-type mice were typically used as recipients for transplantation at an age of 8–10 weeks. Recipient mice received 1,100 cGy of split-dosed lethal irradiation (550 cGy × 2) 3–4 h apart to reduce gastrointestinal toxicity. To obtain LP bone marrow cells from euthanized donor mice, the femurs and tibias were collected aseptically and the bone marrow canals washed out with sterile media. Bone marrow cells were depleted of T cells by incubation with

anti-Thy 1.2 and low-TOX-M rabbit complement (Cedarlane Laboratories). The TCD bone marrow was analysed for purity by quantification of the remaining T cell contamination using flow cytometry. T cell contamination was usually about 0.2% of all leukocytes after a single round of complement depletion. LP donor T cells were prepared by collecting splenocytes aseptically from euthanized donor mice. T cells were purified using positive selection with CD5 magnetic Microbeads with the MACS system (Miltenyi Biotec). T cell purity was determined by flow cytometry, and was routinely approximately 90%. Recipients typically received 5×10^6 TCD bone marrow cells with or without 4×10^6 T cells per mouse via tail vein injection.

Mice were monitored daily for survival and weekly for GVHD scores with an established clinical GVHD scoring system (including weight, posture, activity, fur ruffling and skin integrity) as previously described³⁸. A clinical GVHD index with a maximum possible score of ten was then generated. Mice with a score of five or greater were considered moribund and euthanized by CO₂ asphyxia.

In vivo cytokine administration. Recombinant mouse IL-22 was purchased from GenScript and reconstituted as described by the manufacturer to a concentration of $40 \mu\text{g ml}^{-1}$ in PBS. Mice were treated daily via i.p. injection with either $100 \mu\text{l}$ PBS or $100 \mu\text{l}$ PBS containing $4 \mu\text{g ml}^{-1}$ IL-22. IL-22 administration was started on day 7 after BMT. This schedule was based on the results of rmlIL-22 pharmacokinetics tested in untransplanted mice. For *in vivo* F-652 administration, starting from day 7 after BMT, mice were injected subcutaneously every other day for ten consecutive weeks with PBS or $100 \mu\text{g kg}^{-1}$ F-652.

Histopathology analysis of GVHD target organs. Mice were euthanized for organ analysis 21 days after BMT using CO₂ asphyxiation. For histopathological analysis of GVHD, the small and large intestines were formalin-preserved, paraffin-embedded, sectioned and stained with haematoxylin and eosin. An expert in the field of GVHD pathology, blinded to allocation, assessed the sections for markers of GVHD histopathology. As described previously³⁸, a semiquantitative score consisting of 19 different parameters associated with GVHD was calculated.

LacZ staining. For evaluation of stem-cell numbers, small intestines from Lgr5-LacZ recipient mice that were transplanted with LP bone marrow (and T cells where applicable) were collected. β -galactosidase (LacZ) staining was performed as previously described previously¹. Washed 2.5-cm-sized small intestine fragments were incubated with an ice-cold fixative, consisting of 1% formaldehyde, 0.2% NP40 and 0.2% glutaraldehyde. After removing the fixative, organs were stained for the presence of LacZ according to manufacturer's protocol (LacZ staining kit, Invivogen). The organs were then formalin-preserved, paraffin-embedded, sectioned and counterstained with Nuclear Fast Red (Vector Labs).

Immunohistochemistry staining. Immunohistochemistry detection of REG3 β was performed at the Molecular Cytology Core Facility of MSKCC using a Discovery XT processor (Ventana Medical Systems). Formalin-fixed tissue sections were deparaffinized with EZPrep buffer (Ventana Medical Systems), antigen retrieval was performed with CC1 buffer (Ventana Medical Systems) and sections were blocked for 30 min with Background Buster solution (Innovex). Slides were incubated with anti-REG3 β antibodies (R&D Systems, MAB5110; $1 \mu\text{g ml}^{-1}$) or isotype ($5 \mu\text{g ml}^{-1}$) for 6 h, followed by a 60-min incubation with biotinylated goat anti-rat IgG (Vector Laboratories, PK-4004) at a 1:200 dilution. The detection was performed with a DAB detection kit (Ventana Medical Systems) according to the manufacturer's instructions. Slides were counterstained with haematoxylin (Ventana Medical Systems), and coverslips were added with Permount (Fisher Scientific). See Supplementary Table 1 for full description of antibodies used.

Immunofluorescent staining and microscopic imaging. Immunofluorescent staining was performed at the Molecular Cytology Core Facility of Memorial Sloan Kettering Cancer Center using a Discovery XT processor (Ventana Medical Systems). Formalin-fixed tissue sections were deparaffinized with EZPrep buffer (Ventana Medical Systems), and antigen retrieval was performed with CC1 buffer (Ventana Medical Systems). Sections were blocked for 30 min with Background Buster solution (Innovex) followed by avidin/biotin blocking for 12 min. IL-22R antibodies (R&D Systems, MAB42; $0.1 \mu\text{g ml}^{-1}$) were applied and sections were incubated for 5 h followed by 60 min incubation with biotinylated goat anti-rat IgG (Vector Laboratories, PK-4004) at a 1:200 dilution. The detection was performed with streptavidin–horseradish peroxidase (HRP) D (part of DABMap kit, Ventana Medical Systems), followed by incubation with Tyramide Alexa Fluor 488 (Invitrogen, T20932) prepared according to manufacturer's instruction with predetermined dilutions. Next, lysozyme antibodies (DAKO, A099; $2 \mu\text{g ml}^{-1}$) were applied and sections were incubated for 6 h followed by incubation with biotinylated goat anti-rabbit IgG (Vector Laboratories, PK6101) for 60 min. The detection was performed with streptavidin–HRP D (part of DABMap kit, Ventana Medical Systems), followed by incubation with Tyramide Alexa Fluor 594 (Invitrogen, T20935) prepared according to manufacturer's instruction with predetermined dilutions. Finally, GFP antibodies were applied and sections were

incubated for 5 h followed by incubation with biotinylated goat anti-chicken IgG (Vector Laboratories, BA-9010) for 60 min. The detection was performed with streptavidin–HRP D (part of DABMap kit, Ventana Medical Systems), followed by incubation with Tyramide Alexa Fluor 647 (Invitrogen, T20936) prepared according to manufacturer instruction with predetermined dilutions. Slides were counterstained with DAPI (Sigma Aldrich, D9542; $5 \mu\text{g ml}^{-1}$) for 10 min and coverslips were added with Mowiol. For immunofluorescent and other microscopic imaging, including LacZ and immunohistochemistry slides, contrast and white balance were set based on control slides for each experiment, and the same settings were used for all slides to maximize sharpness and contrast. See Supplementary Table 1 for full description of antibodies used.

Cytokine multiplex assay. Spleen and small intestine were collected from euthanized BMT recipients, and organs were then homogenized and spun down. The supernatant was stored at -20°C until use for cytokine analysis. The cytokine multiplex assays were performed on thawed samples with the mouse Th1/Th2/Th17/Th22 13plex (FlowCytomix Multiplex kit, eBioscience) and performed according to the manufacturer's protocol.

Flow cytometry. For *in vivo* experiments, lymphoid organs were collected from euthanized mice and processed into single cell suspension. Cells were stained with the appropriate mixture of antibodies. For intracellular analysis, an eBioscience Fixation/Permeabilization kit was used per the manufacturer's protocol. After thorough washing, the cells were stained with intracellular and extracellular antibodies simultaneously. Fluorochrome-labelled antibodies were purchased from BD Pharmingen (CD4, CD8, CD24, CD25, CD45, $\alpha\beta$ 7 and P-STAT3 Y705, P-STAT1 Y701), eBioscience (FOXP3), R&D (IL-22R), and Invitrogen (GFP). DAPI and Fixable Live/Dead Cell Stain Kits (Invitrogen) were used for viability staining. Paneth cells were identified based on bright CD24 staining and side scatter granularity as described previously².

For flow cytometry of small intestine organoid cells, organoids were dissociated using TrypLE (37°C). After vigorously pipetting through a p200 pipette causing mechanical disruption, the crypt suspension was washed with 10 ml of DMEM/F12 medium containing 10% FBS and 0.8 kU ml^{-1} DNase1 and passed through a cell strainer. Where applicable, the cells were directly stained or first fixed (4% paraformaldehyde) and permeabilized (methanol) depending on the extracellular or intracellular location of the target protein. All stainings with live cells were performed in PBS without Mg^{2+} and Ca^{2+} with 0.5% BSA. For EdU incorporation experiments there was a 1 h pre-incubation of EdU in the ENR medium of the intact organoid cultures before dissociating the cells with TrypLE. Cells were stained using Click-it kits for imaging and flow cytometry (Life Technologies). For cell cycle analysis, single cell suspensions obtained from dissociated organoids were fixed and stained with Hoechst 33342 (Life Technologies), then assessed with flow cytometry for DNA content and ploidy.

For intracellular pSTAT staining of organoids, organoids were mechanically disrupted into crypt fragments, stimulated for 20 min with 20 ng ml^{-1} IL-22 at 37°C , and then fixed with 4% paraformaldehyde (10 min at 37°C). To assess STAT activation in Lgr5⁺ cells, after freshly isolating crypts from Lgr5–GFP mice, single-cell suspensions including Y-27632 ($10 \mu\text{M}$) were stimulated with IL-22. After obtaining a single cell suspension of stimulated and fixed cells, the samples were filtered ($40 \mu\text{M}$) and permeabilized with ice-cold (-20°C) methanol. Fixed and permeabilized cells were rehydrated with PBS and thoroughly washed with PBS before staining, then stained with anti-phospho-STAT3 and anti-phospho-STAT1, plus anti-GFP or cell surface markers, for 30 min at 4°C .

All flow cytometry was performed with an LSRII cytometer (BD Biosciences) using FACSDiva (BD Biosciences), and the data were analysed with FlowJo software (Treestar). See Supplementary Table 1 for full description of antibodies used.

Immunoblotting analysis. Western blot analysis was carried out on total protein extracts. Free-floating crypts isolated from small intestine were treated in DMEM supplemented with Y-27632 (10 ng ml^{-1} , Tocris) and IL-22 (5 ng ml^{-1} , 30 min). Vehicle (PBS) was added to control wells. Crypts were then lysed in RIPA buffer containing a cocktail of protease and phosphatase inhibitors (Sigma). After sonication, protein amount was determined using the bicinchoninic acid assay Kit (Pierce). Loading $30 \mu\text{g}$ per lane of lysate, proteins were separated using electrophoresis in a 10% polyacrylamide gel and transferred to nitrocellulose. Membranes were blocked for 1 h at room temperature with 1% Blot-Qualified BSA (Promega, W384A) and 1% non-fat milk (LabScientific, M0841) and then incubated overnight at 4°C with the following primary antibodies: rabbit anti-phospho-STAT1 (7649P), rabbit anti-phospho-STAT3 (9131S), rabbit anti-STAT1 (9172P) and rabbit anti-STAT3 (4904P), all from Cell Signaling. This was followed by incubation with the secondary antibody anti-rabbit HRP (7074P2) and visualization with the Pierce ECL Western Blotting Substrate (Thermo Scientific, 32106).

MTT assay. Cell viability in organoids was assessed with a 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium (MTT) test, based on the identification of

metabolically active cells. The organoids were incubated with MTT (0.9 mg ml⁻¹ final concentration, Sigma) for 2 h at 37°C. Matrigel and cells containing intracellular reduction end product formazan were solubilized with acidic isopropanol (isopropanol with HCl) and the reduction end formazan production was evaluated by spectrophotometry using the Infinite M1000 pro plate reader (Tecan).

RT-qPCR. For qPCR, segments of small intestine or isolated crypts were collected from euthanized mice and stored at -80°C. Alternatively, RNA was isolated from organoids after *in vitro* culture. Extracted RNA was also stored at -80°C. Reverse transcriptase PCR (RT-PCR) was performed with a QuantiTect Reverse Transcription Kit (QIAGEN) or a High-Capacity RNA-to-cDNA Kit (Applied Biosystems). qPCR was performed on a Step-One Plus or QuantStudio 7 Flex System (Applied Biosystems) using TaqMan Universal PCR Master Mix (Applied Biosystems). Specific primers were obtained from Applied Biosystems: *Actb*: Mm01205647_g1; *Hprt*: Mm00446968_m1; *Reg3b*: Mm00440616_g1; *Reg3g*: Mm00441127_m1; *Wnt3*: Mm00437336_m1; *Egf*: Mm00438696_m1; *Rspo3*: Mm00661105_m1; *Axin2*: Mm00443610_m1; *Cttnb1*: Mm00483039_m1; *Defa1*: Mm02524428_g1; and *Il22ra1*: Mm01192943_m1. Other primers were obtained from PrimerBank: *Gapdh* (ID 6679937a1), *Cdkn1a* (also known as *p21*) (ID 6671726a1), *Cdkn2d* (also known as *p19*) (ID 31981844a1), *Wnt3a* (ID 7106447a1), *Axin2* (ID 31982733a1), *Hes1* (ID 6680205a1), *Dll4* (ID 9506547a1), *Dll1* (ID 6681197a1), for which cDNAs were amplified with SYBR master mix (Applied Biosystems) in QuantStudio 7 Flex System (Applied Biosystems). Relative amounts of mRNA were calculated by the comparative ΔC_t method with *Actb*, *Hprt* or *Gapdh* as house-keeping genes.

For *Il22ra1* qPCR on Lgr5⁺ cells, dissociated crypt cells from Lgr5-GFP mice were stained and isolated using the following monoclonal antibodies/parameters: EpCAM-1 (G8.8; BD Bioscience); CD45 (30F11; Life Technologies); CD31 (390; BioLegend), Ter119 (Ter119; BioLegend); GFP expression; dead cells were excluded using 7AAD. Cells were acquired on a BD ARIAIII and FACS-sorted. Cells were sorted directly into RA-1/TCEP (Macherey-Nagel) lysis buffer and stored at -80°C until further analysis. RNA of haematopoietic cells (composite of dendritic cells, ILCs and B cells) was used as negative control. RNA was extracted using the NucleoSpin RNA XS kit (Machery Nagel) and cDNA was prepared with Ovation Pico and PicoSL WTA Systems V2 (NuGen). For qPCR, a Nevi Thermal Cycler (Applied Biosystems) and DyNamo Flash SYBR Green qPCR kit (Finnzymes) were used, with the addition of MgCl₂ to a final concentration of 4 mM. All reactions were done in duplicate and normalized to *Gapdh*. Relative expression was calculated by the cycling threshold (C_t) method as $2^{-\Delta C_t}$. The primer sequences were as follows: *Il22ra1*: forward 5'-TCGGCTTGCTCTGTTATC-3', reverse 5'-CCACTGAGGTCCAAGACA-3'.

GSEA. To explore the association of ISC gene signatures (GSE33948 and GSE23672)¹⁶ with STAT3-regulated genes, we performed GSEA in a mouse DSS colitis data set (GSE15955)¹², comparing *Stat3^{fl/fl}*; *Villin-Cre*⁻ (wild type) and *Stat3^{fl/fl}*; *Villin-Cre*⁺ (*Stat3^{ΔIEC}*) mice with DSS colitis (GSEA2-2.2.0; <http://www.broadinstitute.org/gsea>)^{39,40}. A Paneth cell signature gene set was used as a negative control (*DLL1*⁺*CD24^{hi}*, GSE39915)¹⁷. Nominal *P* values are shown.

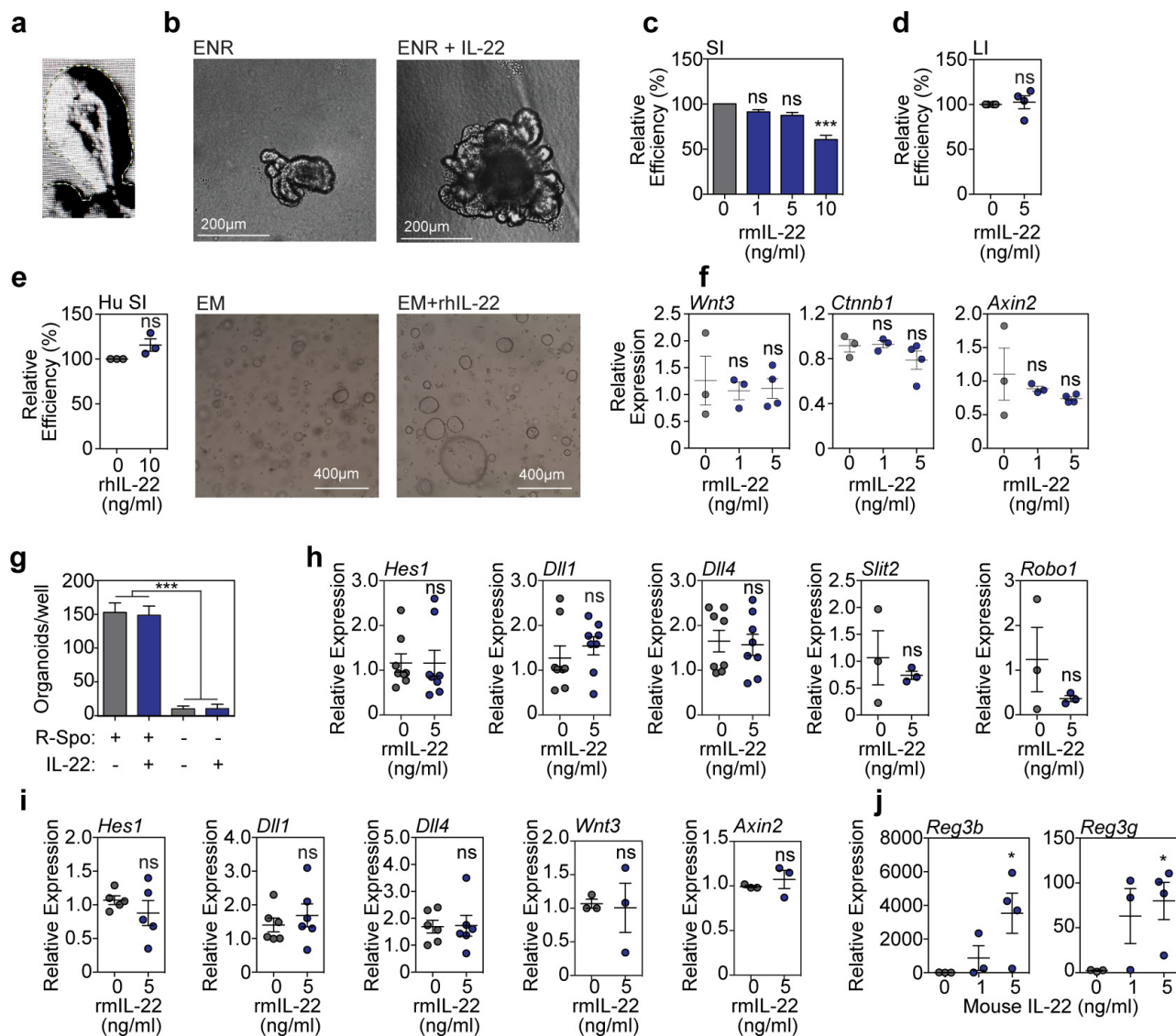
Statistics and software. No statistical methods were used to predetermine sample size. To detect an effect size of >50% difference in means, with an assumed coefficient of variation of 30%, common in biological systems, we attempted to

have at least five samples per group, particularly for *in vivo* studies. All experiments were repeated at least once. No mice were excluded from experiments. Experiments that were technical failures, such as experiments *in vitro* where cultures did not grow or experiments *in vivo* where transplanted control mice (bone marrow plus T cells) did not develop GVHD, were not included for analysis. Occasional individual mice that died post-transplant before analysis could not be included for tissue evaluation.

All data are mean and s.e.m. for the various groups. Statistics are based on 'n' biological replicates. All tests performed are two sided. For the comparisons of two groups, a *t*-test or non-parametric test was performed. Adjustments for multiple comparisons were made. In most cases, non-parametric testing was performed if normal distribution could not be assumed. RT-qPCR reactions and ordinal outcome variables were tested non-parametrically. All analyses of statistical significance were calculated and displayed compared with the reference control group unless otherwise stated.

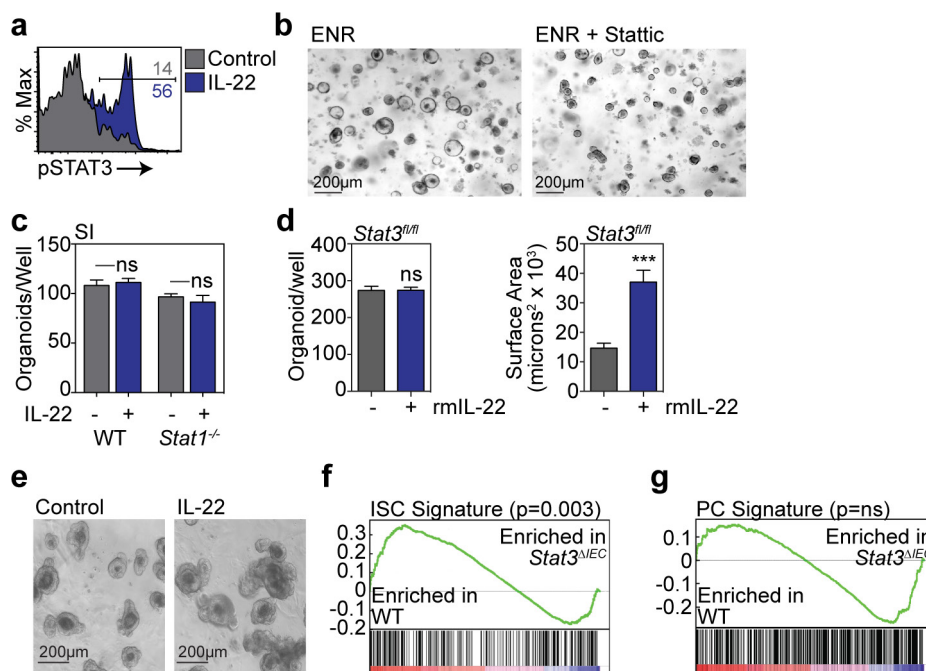
There is large biological variation in organoid size. Statistical analyses of organoid sizes were thus based on all evaluable organoids (at least 25 organoids per group for all experiments). Statistical analyses of organoid numbers and efficiency were based on individual wells. To take into account intra-individual and intra-experimental variation as well, all *in vitro* experiments were performed at least twice with several wells per condition, and sample material coming from at least two different mice. Statistical analyses of stem-cell numbers (Lgr5-LacZ mice) *in vivo* were performed on several independent sections from multiple mice. Statistics were calculated and display graphs were generated using Graphpad Prism.

31. Ootani, A. *et al.* Sustained *in vitro* intestinal epithelial culture within a Wnt-dependent stem cell niche. *Nature Med.* **15**, 701–706 (2009).
32. Wang, F. *et al.* Isolation and characterization of intestinal stem cells based on surface marker combinations and colony-formation assay. *Gastroenterology* **145**, 383–395 (2013).
33. Magness, S. T. *et al.* A multicenter study to standardize reporting and analyses of fluorescence-activated cell-sorted murine intestinal epithelial cells. *Am. J. Physiol. Gastrointest. Liver Physiol.* **305**, G542–G551 (2013).
34. Spits, H. *et al.* Innate lymphoid cells—a proposal for uniform nomenclature. *Nature Rev. Immunol.* **13**, 145–149 (2013).
35. Zheng, Y. *et al.* Interleukin-22 mediates early host defense against attaching and effacing bacterial pathogens. *Nature Med.* **14**, 282–289 (2008).
36. Shroyer, N. F. *et al.* Intestine-specific ablation of mouse atonal homolog 1 (*Math1*) reveals a role in cellular homeostasis. *Gastroenterology* **132**, 2478–2488 (2007).
37. Alpdogan, Ö. *et al.* IL-7 enhances peripheral T cell reconstitution after allogeneic hematopoietic stem cell transplantation. *J. Clin. Invest.* **112**, 1095–1107 (2003).
38. Cooke, K. R. *et al.* An experimental model of idiopathic pneumonia syndrome after bone marrow transplantation: I. The roles of minor H antigens and endotoxin. *Blood* **88**, 3230–3239 (1996).
39. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
40. Mootha, V. K. *et al.* PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genet.* **34**, 267–273 (2003).



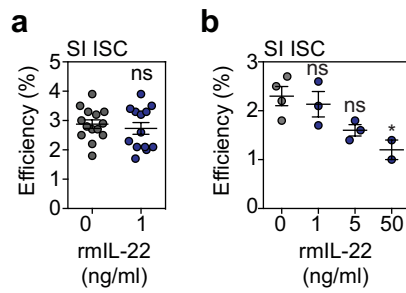
Extended Data Figure 1 | IL-22 increases organoid growth without activating the Wnt or Notch pathways. **a**, Microscopic tracing of organoid to measure surface area. **b**, Brightfield images of SI organoids from B6 mice, after 7 days of culture with/without IL-22 (5 ng ml⁻¹). **c–e**, Organoid efficiency (percentage) relative to control (0 ng ml⁻¹) for B6 SI organoids (statistics on data combined from $n = 19$ wells per group from 19 individual mice) (**c**), B6 large intestine organoids ($n = 4$ mice per group) cultured with/without rmIL-22 for 7 days (**d**), and human SI organoids cultured with/without rhIL-22 for 6 days ($n = 3$ donors per group) (**e**). **f**, RT-qPCR of relative mRNA expression of *Wnt3*, *Ctnnb1* and *Axin2* genes of the Wnt/β-catenin axis in SI organoids cultured with/without rmIL-22; $n = 3$ (0–1 ng ml⁻¹) and $n = 4$ (5 ng ml⁻¹) mice per group. **g**, Numbers of SI organoids per well with/without rmIL-22 (5 ng ml⁻¹) in the presence or absence of R-spondin-1 ($n = 6$ wells per group). **h**, RT-qPCR-determined

relative mRNA expression of Notch pathway genes (*Hes1*, *Dll1* and *Dll4*; $n = 8$ mice per group) as well as of *Slit2* and its receptor *Robo1* ($n = 3$ mice per group) in day-7 SI organoids cultured with/without rmIL-22. **i**, Relative expression of *Wnt3* and *Axin2* ($n = 3$ mice per group), *Hes1* ($n = 5$ mice per group), and *Dll1* and *Dll4* ($n = 6$ mice per group) genes in large intestine organoids. **j**, RT-qPCR for the relative mRNA expression of *Reg3b* and *Reg3g* innate antimicrobials in SI organoids cultured with rmIL-22; $n = 3$ (0–1 ng ml⁻¹) and $n = 4$ (5 ng ml⁻¹) mice per group. Organoid efficiency and number comparisons were performed with *t*-tests (two groups) or ANOVA (multiple groups). RT-qPCR statistics were performed with non-parametric Mann–Whitney *U* (two groups) or Kruskal–Wallis (multiple groups) tests. Data are mean and s.e.m.; * $P < 0.05$, *** $P < 0.001$. Data combined from at least two independent experiments unless otherwise stated.

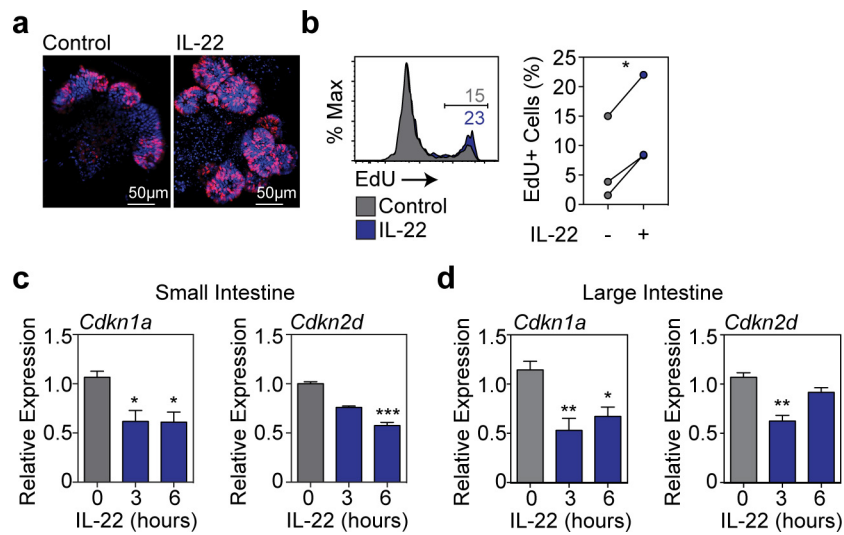


Extended Data Figure 2 | IL-22 activates STAT3 in intestinal organoids, and STAT3 deficiency leads to ISC gene signature loss in mice with colitis. **a**, Intracellular staining of pSTAT3 (Y705) in organoid cells cultured under standard ENR conditions followed by a 20 min pulse of 20 ng ml⁻¹ IL-22, evaluated by flow cytometry; data representative of two independent experiments. **b**, Brightfield images of SI organoids 4 days after crypt culture with/without Stattic; data representative of three experiments. **c**, SI organoids per well from wild-type and *Stat1*^{-/-} mice with/without rmIL-22; *n* = 6 wells per group; ANOVA. **d**, **e**, Day 5 organoids from *Stat3*^{fl/fl} SI crypt cells cultured with/without rmIL-22 (5 ng ml⁻¹) in the absence of adeno-Cre infection; numbers per well

(*n* = 6 wells per group) and size (*n* = 35 control and *n* = 42 IL-22-treated organoids per group), *t*-test (**d**); brightfield images representative of three experiments (**e**). **f**, **g**, GSEAs of the expression of a second independent ISC signature gene set (GSE36497) (**f**) and a negative control DLL1⁺CD24^{hi} Paneth cell (PC) gene set (GSE39915) (**g**) in *Stat3*^{fl/fl}; *Villin-Cre*⁻ (wild type) versus *Stat3*^{fl/fl}; *Villin-Cre*⁺ (*Stat3*^{ΔIEC}) mice with DSS colitis, using GEO database array data (GSE15955). Each GSEA represents one analysis; nominal *P* values are shown. Data are mean and s.e.m.; ****P* < 0.001. Data combined from at least two independent experiments unless otherwise stated.

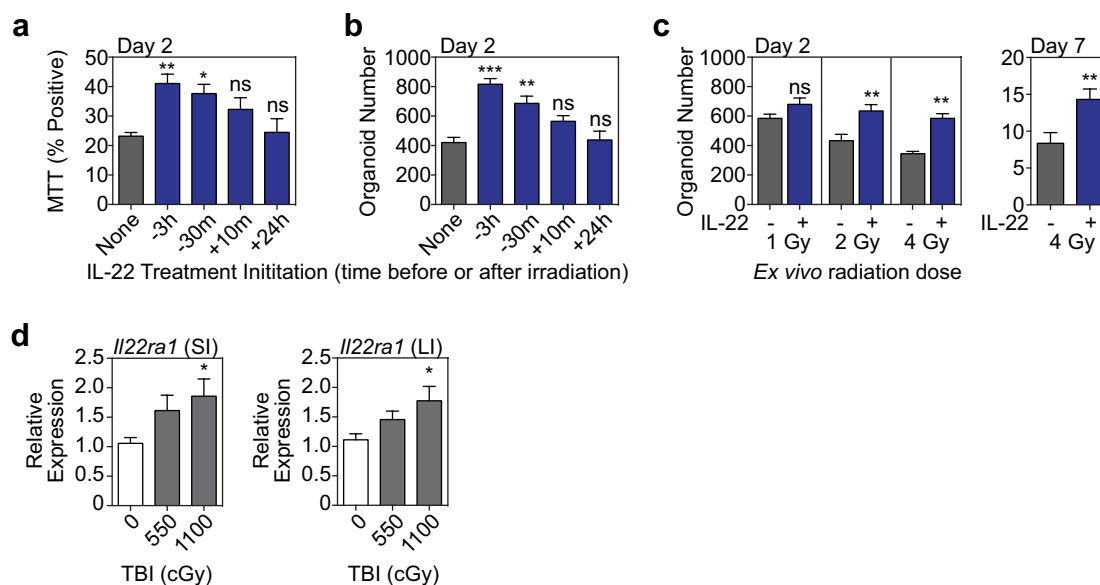


Extended Data Figure 3 | Efficiency of organoid formation from purified ISCs cultured with IL-22. **a, b**, Organoid efficiency as percentage of plated cells, in organoid cultures from sorted Lgr5⁺ ISCs from B6 Lgr5-GFP reporter mice using a concentration of 1 ng ml⁻¹ ($n = 14$ wells per group combined from three experiments; t -test) (**a**) and with a concentration range (one experiment, $n = 3$ wells per group; ANOVA) (**b**). Data are mean and s.e.m.; * $P < 0.05$.



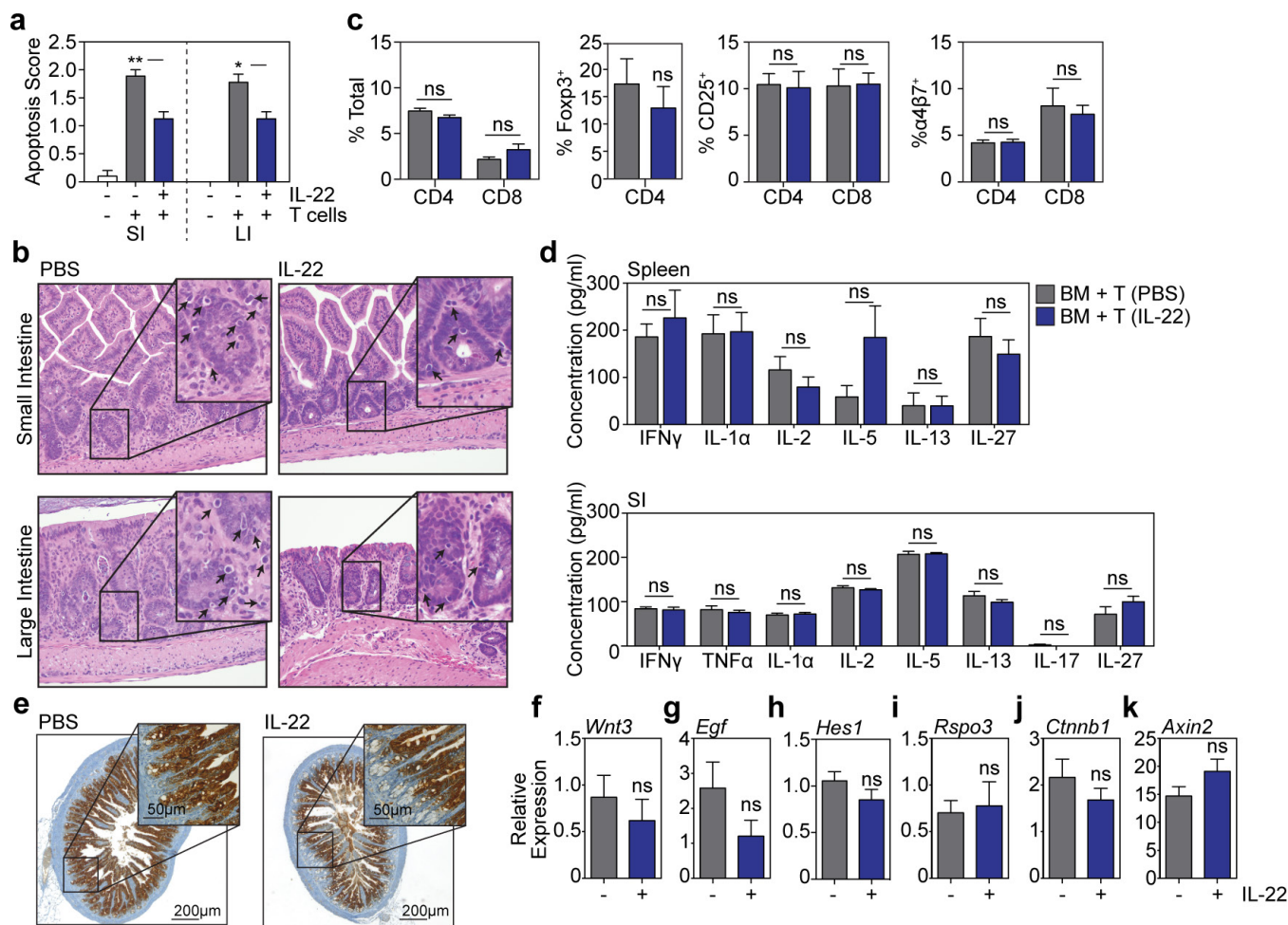
Extended Data Figure 4 | IL-22 increases cellular proliferation in intestinal organoids. **a, b**, Confocal images (nuclear staining, blue; and EdU staining, red; one experiment) (**a**) and FACS analysis (**b**) of EdU incorporation (1 h) in SI organoids cultured in the presence or absence of rmIL-22 (1 ng ml⁻¹); histogram representative of two experiments, graph shows paired *t*-test, *n* = 3 mice per group combined from two experiments.

c, d, *Cdkn1a* and *Cdkn2d* mRNA expression (RT-qPCR) in organoids cultured from small (**c**) and large (**d**) intestine crypts for 24 h with 0, 3 or 6 h exposure to IL-22 before collection; Kruskal–Wallis analysis, *n* = 6 mice per group combined from two independent experiments. Data are mean and s.e.m.; **P* < 0.05, ***P* < 0.01, ****P* < 0.001.



Extended Data Figure 5 | Intestinal organoids and crypts after irradiation. **a–c**, Dissociated single cells from wild-type B6 crypts were exposed to escalating doses of irradiation *ex vivo*. **a**, **b**, Crypt cells were plated 3 h before irradiation, and cultures were treated with rmIL-22 (5 ng ml⁻¹) added to the culture at 3 h before, 30 min before, 10 min after or 24 h after 4 Gy irradiation. Two days after irradiation, organoids were evaluated for MTT viability testing (percentage positive, $n = 6$ wells per group) (**a**) and the number of organoids generated ($n = 6$ wells per group) (**b**). **c**, The effect of IL-22 after irradiation was evaluated by measuring number of organoids 2 days and 7 days after irradiation (day 2:

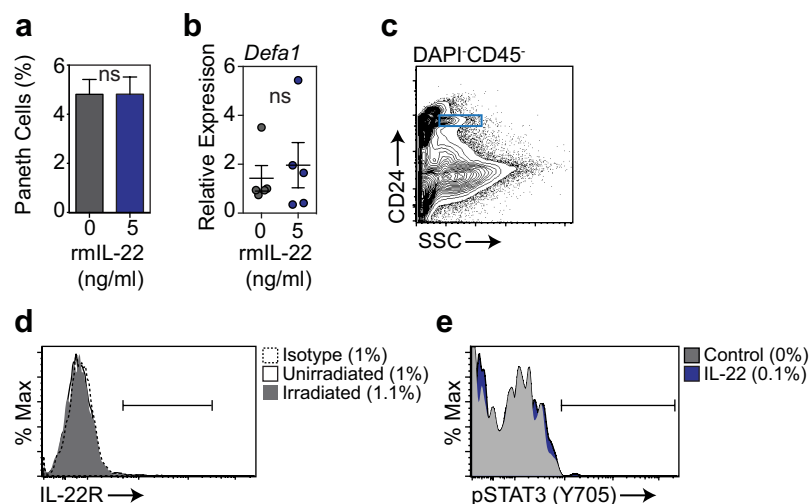
$n = 9$ wells per group for 1–2 Gy and $n = 6$ wells per group for 4 Gy; day 7: 4 Gy, $n = 20$ wells per group). Culture with/without IL-22 was initiated 3 h before irradiation. **d**, Small and large intestine crypt *Il22ra1* expression determined by qPCR; RNA isolated from fresh crypts of B6 mice collected 1 day (20–26 h) after total body irradiation; $n = 12$ control and $n = 11$ irradiated mice per group. Comparisons performed with *t*-tests (two groups) or ANOVA (multiple groups). Data are mean and s.e.m.; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Data combined from at least two independent experiments.



Extended Data Figure 6 | IL-22 treatment after allogeneic BMT.

B6 recipient mice were transplanted with only TCD bone marrow from LP donors, or with bone marrow and T cells from LP donors to induce GVHD (H-2^b into H-2^b). Mice receiving T cells were treated daily with PBS or 4 μ g rmIL-22 by i.p. injection starting 7 days after BMT. **a**, Pathological scoring of apoptosis in intestinal tissues 3 weeks after BMT. Data from two experiments combined; $n = 10$ (TCD bone marrow only mice), $n = 9$ (BM + T (PBS)), $n = 8$ (BM + T (IL-22)); Kruskal–Wallis analysis. **b**, Representative haematoxylin and eosin staining of small and large intestines. Arrows indicate apoptotic cells within the intestinal epithelium. **c**, Splenocytes from recipients were analysed by flow cytometry 3 weeks after BMT, indicating frequencies of T cell subsets, expression of activation marker CD25, and expression of gut homing molecule $\alpha_4\beta_7$ integrin; $n = 9$ (PBS-treated) and $n = 10$ (IL-22-treated) mice per group; t -test analysis.

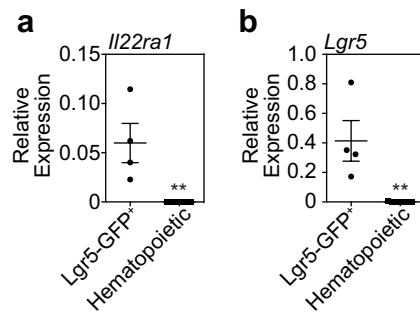
d, Expression of inflammatory cytokines in spleen ($n = 9$ PBS-treated and $n = 10$ IL-22-treated mice per group) and SI ($n = 10$ mice per group) was analysed in recipient tissues 3 weeks after BMT; t -test analyses, multiple comparisons corrected for with Holm–Sidak correction. **e**, REG3 β immunohistochemistry staining in SI samples of recipient mice 3 weeks after BMT, data representative of three experiments. **f–k**, RT–qPCR of relative mRNA expression in SI tissue samples of PBS-treated versus IL-22-treated mice 3 weeks post-BMT for: *Wnt3* (**f**); *Egf* (**g**); *Hes1* (from purified crypts) (**h**); *Rspo3* (**i**); *Ctnnb1* (from purified crypts) (**j**); *Axin2* (from purified crypts) (**k**); $n = 10$ mice per group for purified crypt samples; $n = 8$ (PBS-treated) and $n = 9$ (IL-22-treated) mice per group for whole SI tissue samples; Mann–Whitney U test. Data are mean and s.e.m.; * $P < 0.05$, ** $P < 0.01$. Data combined from two independent experiments unless stated otherwise.



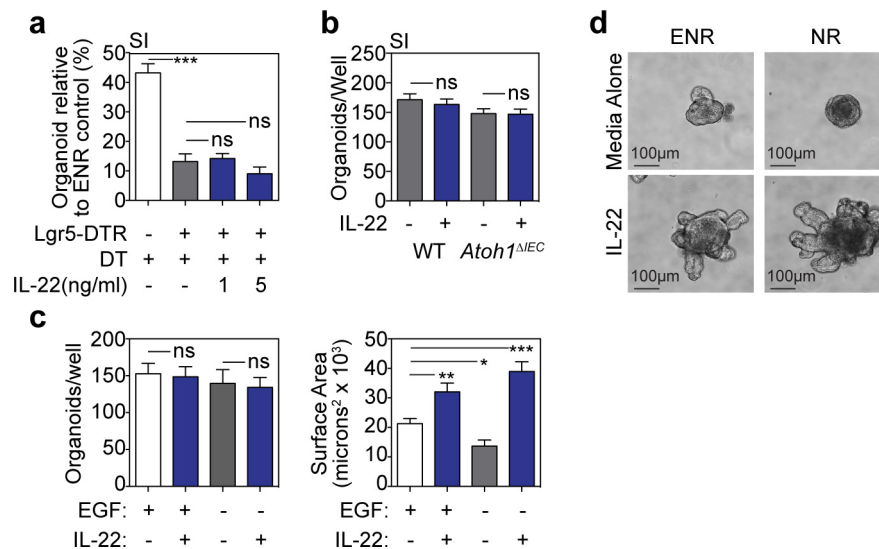
Extended Data Figure 7 | IL-22 does not enhance Paneth cell frequency, *Defa1* gene expression, or STAT3 phosphorylation *in vitro*.

a, Percentage of Paneth cells in organoids cultured with/without 5 ng ml⁻¹ rmIL-22 for 7 days, as evaluated by flow cytometry after dissociation into single cells; *n* = 7 independent cultures per group (one mouse per culture); *t*-test. **b**, RT-qPCR analysis of the relative mRNA expression of Paneth cell gene *Defa1* in SI organoids cultured with/without 5 ng ml⁻¹ rmIL-22 for 7 days; *n* = 5 independent cultures per group (1–2 pooled mice per culture); Mann–Whitney *U* test. **c–e**, Paneth cell IL-22R expression and

STAT3 phosphorylation assessed by flow cytometry. Shown are gating of Paneth cells based on side scatter and CD24 expression (**c**), Paneth cell IL-22R expression at baseline and 5 days after 1,200 cGy total body irradiation (one of two experiments) (**d**), and STAT3 phosphorylation in Paneth cells as determined by phosflow of dissociated crypt cells after a 20-min pulse with rmIL-22 (20 ng ml⁻¹, 37 °C; one of two experiments) (**e**). Data are mean and s.e.m. Data combined from at least four independent experiments unless otherwise stated.

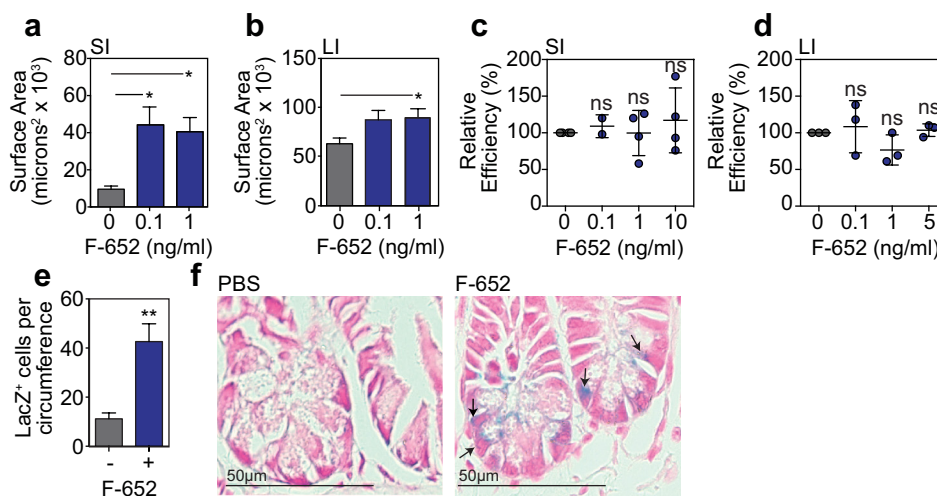


Extended Data Figure 8 | ISCs express *Il22ra1*. **a**, Relative mRNA expression of *Il22ra1* in sorted Lgr5-GFP⁺ cells ($n = 4$ biological replicates), with various sorted haematopoietic populations serving as negative controls, including intestinal dendritic cells ($n = 4$), intestinal ILC3s ($n = 2$), and splenic B cells ($n = 1$). **b**, *Lgr5* mRNA relative to *Gapdh* expression in sorted Lgr5-GFP⁺ cells and haematopoietic samples described above to confirm Lgr5 expression in sorted Lgr5-GFP⁺ cells. Data are mean and s.e.m.; Mann-Whitney U test; ** $P < 0.01$.



Extended Data Figure 9 | IL-22 increases the size of SI organoids cultured without EGF. **a**, Efficiency of wild-type and Lgr5-DTR SI organoid formation after culture with diphtheria toxin ($1 \text{ ng } \mu\text{l}^{-1}$) to deplete Lgr5⁺ cells; one of three experiments; $n = 6$ (wild type), $n = 5$ (Lgr5-DTR), $n = 6$ (1 ng ml^{-1} IL-22), $n = 6$ (5 ng ml^{-1} IL-22) wells per group. **b**, Numbers of wild-type and *Atoh1*^{ΔIEC} day-7 SI organoids cultured with/without rmIL-22 (5 ng ml^{-1}); $n = 6$ wells per group. **c**, **d**, Omission of EGF from the standard ENR medium (NR). **c**, The effect of IL-22 on organoid numbers and size in the absence of EGF; $n = 6$ wells per group

for numbers; $n = 45$ (ENR), $n = 37$ (ENR plus IL-22), $n = 42$ (NR), $n = 54$ (NR plus IL-22) organoids per group for size; data combined from three experiments. **d**, Brightfield images of wild-type SI organoid cultures in the presence or absence of EGF (50 ng ml^{-1}), representative of three experiments. Data are mean and s.e.m. Comparisons were performed with *t*-tests (two groups) or ANOVA (multiple groups); * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Data combined from three independent experiments unless otherwise stated.



Extended Data Figure 10 | F-652 increases organoid size *ex vivo* and reduces radiation injury to the ISC compartment *in vivo*. **a, b**, Area of small (**a**) and large (**b**) intestine wild-type B6 organoids cultured with/without the rhIL-22-dimer and Fc-fusion molecule F-652; SI: $n = 37$ (0 ng ml⁻¹), $n = 60$ (0.1 ng ml⁻¹), and $n = 41$ (1 ng ml⁻¹) organoids per group combined from three experiments; LI: $n = 137$ (0 ng ml⁻¹), $n = 83$ (0.1 ng ml⁻¹) and $n = 132$ (1 ng ml⁻¹) organoids per group combined from two experiments; ANOVA. **c, d**, Organoid efficiency relative to control in cultures of B6 SI organoids ($n = 4$ wells per group combined from two experiments) (**c**) and B6 LI organoids ($n = 3$ wells per group; one of two experiments) (**d**) treated with different concentrations

of recombinant human F-652; ANOVA. **e, f**, B6 Lgr5-LacZ mice were treated with PBS or F-652 (100 μg kg⁻¹), administered subcutaneously on the day of total body irradiation (10–12 Gy) and again 2 days later; one of three experiments. **e**, Lgr5-LacZ⁺ crypt cells per SI circumference were evaluated at day 3.5 after irradiation (10 Gy); statistics based on $n = 11$ independent sections (PBS-treated) versus $n = 14$ independent sections (F-652-treated) from irradiated mice; independent sections were derived from three mice per group; first dose of PBS or F-652 was administered 4 h before irradiation; Mann–Whitney *U* test. **f**, Representative crypt base images 3.5 days after irradiation (10 Gy). Arrows indicate Lgr5-LacZ⁺ crypt cells. Data are mean and s.e.m.; * $P < 0.05$, ** $P < 0.01$.

Unique role for ATG5 in neutrophil-mediated immunopathology during *M. tuberculosis* infection

Jacqueline M. Kimmey¹, Jeremy P. Huynh¹, Leslie A. Weiss¹, Sunmin Park², Amal Kambal², Jayanta Debnath³, Herbert W. Virgin² & Christina L. Stallings¹

Mycobacterium tuberculosis, a major global health threat, replicates in macrophages in part by inhibiting phagosome-lysosome fusion, until interferon- γ (IFN γ) activates the macrophage to traffic *M. tuberculosis* to the lysosome. How IFN γ elicits this effect is unknown, but many studies suggest a role for macroautophagy (herein termed autophagy), a process by which cytoplasmic contents are targeted for lysosomal degradation¹. The involvement of autophagy has been defined based on studies in cultured cells where *M. tuberculosis* co-localizes with autophagy factors ATG5, ATG12, ATG16L1, p62, NDP52, BECN1 and LC3 (refs 2–6), stimulation of autophagy increases bacterial killing^{6–8}, and inhibition of autophagy increases bacterial survival^{1,2,4,6,7}. Notably, these studies reveal modest (~ 1.5 –3-fold change) effects on *M. tuberculosis* replication. By contrast, mice lacking ATG5 in monocyte-derived cells and neutrophils (polymorphonuclear cells, PMNs) succumb to *M. tuberculosis* within 30 days^{4,9}, an extremely severe phenotype similar to mice lacking IFN γ signalling^{10,11}. Importantly, ATG5 is the only autophagy factor that has been studied during *M. tuberculosis* infection *in vivo* and autophagy-independent functions of ATG5 have been described^{12–18}. For this reason, we used a genetic approach to elucidate the role for multiple autophagy-related genes and the requirement for autophagy in resistance to *M. tuberculosis* infection *in vivo*. Here we show that, contrary to expectation, autophagic capacity does not correlate with the outcome of *M. tuberculosis* infection. Instead, ATG5 plays a unique role in protection against *M. tuberculosis* by preventing PMN-mediated immunopathology. Furthermore, while *Atg5* is dispensable in alveolar macrophages during *M. tuberculosis* infection, loss of *Atg5* in PMNs can sensitize mice to *M. tuberculosis*. These findings shift our understanding of the role of ATG5 during *M. tuberculosis* infection, reveal new outcomes of ATG5 activity, and shed light on early events in innate immunity that are required to regulate disease pathology and bacterial replication.

We first replicated the finding that *Atg5* is critical in myeloid-derived cells for resistance to *M. tuberculosis* by infecting *Atg5^{fl/fl}-Lysm-cre* (*Lysm* is also known as *Lyz2*) mice^{4,9}. *Lysm*-promoter-driven expression of Cre recombinase (*Lysm-cre*) results in deletion of a floxed gene in alveolar macrophages, recruited macrophages, inflammatory monocytes, monocyte-derived dendritic cells, and PMNs^{19,20}. Following aerosol inoculation of *M. tuberculosis* into wild-type C57Bl/6 mice, bacteria replicate in innate immune cells until IFN γ -producing T cells are recruited to the lungs between 18–20 days post-infection (d.p.i.), resulting in control of bacterial burden and survival²¹. Consistent with previous publications^{4,9}, *Atg5^{fl/fl}-Lysm-cre* mice lost 23% of their weight by 20 d.p.i. and succumbed to *M. tuberculosis* between 30 and 40 d.p.i. (Fig. 1a, b). In contrast, *Atg5^{fl/fl}* control mice showed no signs of sickness or weight loss. Bacterial titres in *Atg5^{fl/fl}-Lysm-cre* mice were significantly higher at 3 weeks post-infection (w.p.i.) than those in

Atg5^{fl/fl} mice (Fig. 1c, d). By 5 w.p.i., *Atg5^{fl/fl}* mice had controlled pulmonary burden while *Atg5^{fl/fl}-Lysm-cre* mice rapidly succumbed to infection (Fig. 1b, c).

In cultured cells, *Atg5*, *p62* (also known as *Sqstm1*) and *Ulk1* have similar roles in controlling *M. tuberculosis* survival and replication^{1,4,5,22}. We therefore explored the role of these and other genes involved in autophagy *in vivo*, by infecting mice with germline deletions of *Ulk1*, *Ulk2* (autophagy induction), *Atg4b* (isolation membrane elongation), or *p62* (substrate targeting to autophagosome). Surprisingly, mice lacking *Ulk1*, *Ulk2*, *Atg4b* or *p62* showed no signs of sickness during infection, efficiently controlled bacterial burden, and survived over 80 days with *M. tuberculosis* (Fig. 1e–h, and Extended Data Fig. 1a). Potential redundancy may explain the lack of a phenotype in *Ulk1^{-/-}*, *Ulk2^{-/-}*, *Atg4b^{-/-}*, and *p62^{-/-}* mice during *M. tuberculosis* infection. However, loss of either *Ulk1* or *Ulk2* results in clear autophagy defects in cultured cells²³, and *Atg4b^{-/-}* mice have dramatic autophagy defects in many tissues, including a nearly complete loss of LC3-II (the lipidated form of microtubule-associated protein 1A/1B-light chain 3 (LC3), which localizes to autophagosome membranes) formation in the lungs, kidney and liver²⁴. Regardless of issues with redundancy, these data indicate a lack of correlation between *in vitro* and *in vivo* findings of the roles of these genes in controlling *M. tuberculosis* replication.

We next tested the role of essential *Atg* genes other than *Atg5* in resistance to *M. tuberculosis*. If ATG5 is required *in vivo* due to its role in canonical autophagy, then *Lysm-cre* deletion of other essential autophagy genes would result in a similar phenotype as observed in *Atg5^{fl/fl}-Lysm-cre* mice. Contrary to expectation, *Atg14^{fl/fl}-Lysm-cre*, *Atg12^{fl/fl}-Lysm-cre*, *Atg16l1^{fl/fl}-Lysm-cre*, *Atg7^{fl/fl}-Lysm-cre* and *Atg3^{fl/fl}-Lysm-cre* mice did not show any signs of sickness or weight loss following infection with *M. tuberculosis* and all survived over 80 d.p.i. (Fig. 1i and Extended Data Fig. 1b). In addition, these mice were all able to control *M. tuberculosis* burden in a manner similar to C57Bl/6 mice (Fig. 1j, k). These findings were particularly notable as these same *Atg16l1^{fl/fl}-Lysm-cre*, *Atg7^{fl/fl}-Lysm-cre* and *Atg3^{fl/fl}-Lysm-cre* mice are dramatically more susceptible to *Toxoplasma gondii*, another pathogen for which IFN γ plays a key role in resistance to infection^{12,14}. Nevertheless, to compare the relative efficacy of conditional deletion of each essential autophagy factor, LC3 lipidation and p62 degradation were measured *ex vivo* in peritoneal exudate macrophages (Fig. 1l) and bronchoalveolar lavage macrophages (Extended Data Fig. 2). Consistent with previous publications using these mouse strains^{12,14}, the floxed alleles in *Atg5^{fl/fl}-Lysm-cre*, *Atg16l1^{fl/fl}-Lysm-cre*, *Atg7^{fl/fl}-Lysm-cre* and *Atg3^{fl/fl}-Lysm-cre* mice were effectively targeted *in vivo* resulting in similar increases in the amounts of LC3-I (the non-lipidated form of LC3, which does not participate in autophagy) and p62, which indicate a defect in autophagy. Peritoneal macrophage and bronchoalveolar macrophages from *Atg14^{fl/fl}-Lysm-cre* mice also

¹Department of Molecular Microbiology, Washington University School of Medicine, St Louis, Missouri 63110, USA. ²Department of Pathology and Immunology, Washington University School of Medicine, St Louis, Missouri 63110, USA. ³Department of Pathology and Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, California 94143, USA.

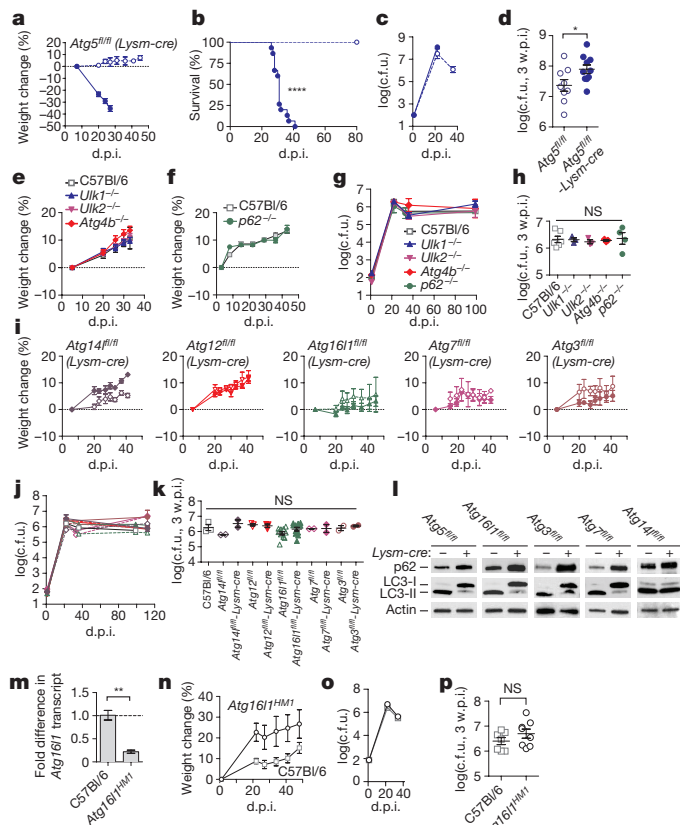


Figure 1 | ATG5, in contrast to other autophagy factors, is essential to control *M. tuberculosis* infection. **a–k**, Mice infected with approximately 100 colony-forming units (c.f.u.) of *M. tuberculosis* were monitored at various days post infection (d.p.i.) or weeks post infection (w.p.i.). **a–d**, Weight change (**a**), survival (**b**), and log pulmonary c.f.u. (**c**, **d**) of *Atg5*^{fl/fl} (open circles) and *Atg5*^{fl/fl}-*Lysm-cre* (closed circles) mice. **e–h**, Weight change (**e**, **f**) and log pulmonary c.f.u. (**g**, **h**) of C57Bl/6 (open squares), *Ulk1*^{-/-} (blue triangles), *Ulk2*^{-/-} (inverted pink triangles), *Atg4b*^{-/-} (red diamonds), and *p62*^{-/-} (green circles) mice. **i–k**, Weight change (**i**) and log pulmonary c.f.u. (**j**, **k**) of *Atg14l1*^{fl/fl}-*Lysm-cre* (purple diamonds), *Atg12l1*^{fl/fl}-*Lysm-cre* (red inverted triangles), *Atg16l1*^{fl/fl}-*Lysm-cre* (green triangles), *Atg7*^{fl/fl}-*Lysm-cre* (pink diamonds), *Atg3*^{fl/fl}-*Lysm-cre* (brown circles) and corresponding floxed control mice. Floxed control mice are shown in open shapes, *LysM*-*Cre*-expressing mice are shown in closed shapes. **l**, Western blot analysis of p62, LC3 and actin in *ex vivo* peritoneal macrophages from uninfected mice. **m**, Fold change in *Atg16l1* transcript from *Atg16l1*^{HMI} lungs as compared to C57Bl/6 at 3 w.p.i. **n–p**, Weight change (**n**) and log pulmonary c.f.u. (**o**, **p**) of *Atg16l1*^{HMI} (open circles) and C57Bl/6 mice (open squares). Statistical differences were determined by log-rank Mantel–Cox test (**b**), Student's *t*-test (**d**, **m** and **p**) or one-way analysis of variance (ANOVA) and Bonferroni's multiple comparison test (**h**, **k**). **P* < 0.05, ***P* < 0.01, *****P* < 0.0001; NS, not significant; error bars represent mean ± s.e.m. Samples represent biological replicates. See Supplementary Fig. 1 for gel source data and Supplementary Fig. 2 for sample sizes and results from all statistical comparisons.

accumulated p62 while, consistent with previous findings, the levels of LC3 were largely unaffected¹².

At 3 w.p.i., *Atg5*^{fl/fl} mice have higher bacterial titres as compared to C57Bl/6 mice (Fig. 1d, k), which we attribute to hypomorphic expression of *Atg5* from the *Atg5*^{fl/fl} allele (Extended Data Fig. 3 and ref. 25). To determine if germline hypomorphism for an essential ATG factor other than ATG5 interferes with control of *M. tuberculosis*, we infected mice that are hypomorphic for ATG16L1 (*Atg16l1*^{HMI})²⁶ (Fig. 1m). *Atg16l1*^{HMI} mice showed no signs of sickness or weight loss following *M. tuberculosis* infection and controlled *M. tuberculosis* burden in a manner similar to C57Bl/6 mice (Fig. 1n–p and Extended Data Fig. 1c). Together, these data demonstrate that the loss of genes essential

for canonical autophagy in *LysM*⁺ cells does not correlate with susceptibility to *M. tuberculosis* and suggest that ATG5 participates in a unique function not served by other essential ATG proteins. While autophagy-independent functions of ATG5 have been described^{12–18}, this is the first example of ATG5 being important for a response to an infection independent of ATG16L1 and ATG12.

To further explore how ATG5 functions during *M. tuberculosis* infection, we next investigated the reports that *Atg5*^{fl/fl}-*Lysm-cre* mice develop more severe inflammation following *M. tuberculosis* infection^{4,9}. Various studies have demonstrated that myeloid-specific defects in components of the membrane elongation complex (ATG5, ATG7 or ATG16L1) can cause increased inflammation *in vivo*^{27–29}. To distinguish between ATG16L1-dependent versus independent roles for ATG5 in regulating inflammation we measured immune responses to *M. tuberculosis* in the lungs of *Atg5*^{fl/fl}-*Lysm-cre*, *Atg16l1*^{fl/fl}-*Lysm-cre* and control mice. Phenotypes specific to loss of *Atg5* might be responsible for susceptibility to *M. tuberculosis* since *Atg16l1*^{fl/fl}-*Lysm-cre* mice control *M. tuberculosis* infection similarly to wild-type C57Bl/6 mice (Fig. 1). At 2 w.p.i., *Atg5*^{fl/fl}-*Lysm-cre* lungs contained larger lesions than those in C57Bl/6, *Atg5*^{fl/fl}, *Atg16l1*^{fl/fl}-*Lysm-cre* and *Atg16l1*^{fl/fl} mice (Fig. 2a), even though bacterial burdens were similar in each strain at this time point (Extended Data Fig. 4). By 3 w.p.i., *Atg5*^{fl/fl}-*Lysm-cre* lungs were severely inflamed with large lesions and extensive consolidation, while *Atg5*^{fl/fl} and *Atg16l1*^{fl/fl}-*Lysm-cre* lungs showed only moderate increases in inflammation (Fig. 2a). Consistent with this, the lungs of *Atg5*^{fl/fl}-*Lysm-cre* mice at 3 w.p.i. contained higher levels of pro-inflammatory cytokines than *Atg16l1*^{fl/fl}-*Lysm-cre* or control mice (Fig. 2b). At this time point, the only cytokine that was significantly higher in the lungs of *Atg16l1*^{fl/fl}-*Lysm-cre* mice compared to controls was IL-1β, however this was still only half as much IL-1β as detected in *Atg5*^{fl/fl}-*Lysm-cre* lungs. The increased levels of IL-1β in mice lacking *Atg16l1* is consistent with previous reports showing that autophagy in macrophages negatively regulates inflammasome-dependent IL-1β production^{9,27–29}. The observed differences in cytokine production were a specific and active response to *M. tuberculosis* infection, as cytokine levels were not significantly different or were below the limit of detection in uninfected lungs (Extended Data Fig. 5).

To characterize cell populations contributing to the inflammation, flow cytometry was performed at 2 and 3 w.p.i. in *Atg5*^{fl/fl}-*Lysm-cre*, *Atg16l1*^{fl/fl}-*Lysm-cre*, and control mice. At 2 w.p.i., *Atg5*^{fl/fl}-*Lysm-cre* lungs contained a significantly greater frequency of PMNs than *Atg5*^{fl/fl} or C57Bl/6 mice (Fig. 2c and Extended Data Fig. 6). This difference was more pronounced at 3 w.p.i., and at this time point the frequency of PMNs in *Atg5*^{fl/fl}-*Lysm-cre* lungs was also significantly higher than in *Atg16l1*^{fl/fl}-*Lysm-cre* lungs (Fig. 2d). *Atg5*^{fl/fl}-*Lysm-cre* lungs also contained a greater percentage of inflammatory monocytes than C57Bl/6 mice at 2 w.p.i., however this level was similar to *Atg5*^{fl/fl} lungs and, by 3 w.p.i., was not significantly different from any other strain. The increased inflammation in *Atg5*^{fl/fl}-*Lysm-cre* lungs likely contributes to the severe lung pathology and morbidity observed in these mice (Fig. 1a, b and Fig. 2a–d). In addition, the absence of higher bacterial burden at 2 w.p.i. (Extended Data Fig. 4) indicates that the increased inflammation in the *M. tuberculosis* infected *Atg5*^{fl/fl}-*Lysm-cre* mice is a direct result of loss of *Atg5* rather than a response to uncontrolled bacterial replication.

Excessive PMN recruitment is a hallmark of acute susceptibility to *M. tuberculosis* and is associated with uncontrolled tissue damage and progression of disease¹¹. We hypothesized that the susceptibility of the *Atg5*^{fl/fl}-*Lysm-cre* mice is related to the increased frequency of PMNs in these mice during *M. tuberculosis* infection and, therefore, sought to determine if depletion of PMNs would improve control of *M. tuberculosis*¹¹. Antibody-mediated depletion of PMNs (anti-Ly6G, clone 1A8) from 10–28 d.p.i. allowed *Atg5*^{fl/fl}-*Lysm-cre* mice to recover their lost weight and survive over 80 d.p.i. (Fig. 3a, b). To survive 80 d.p.i., PMN-depleted *Atg5*^{fl/fl}-*Lysm-cre* mice must have functional IFNγ signalling and T-cell responses, since *Rag*^{-/-} and PMN-depleted

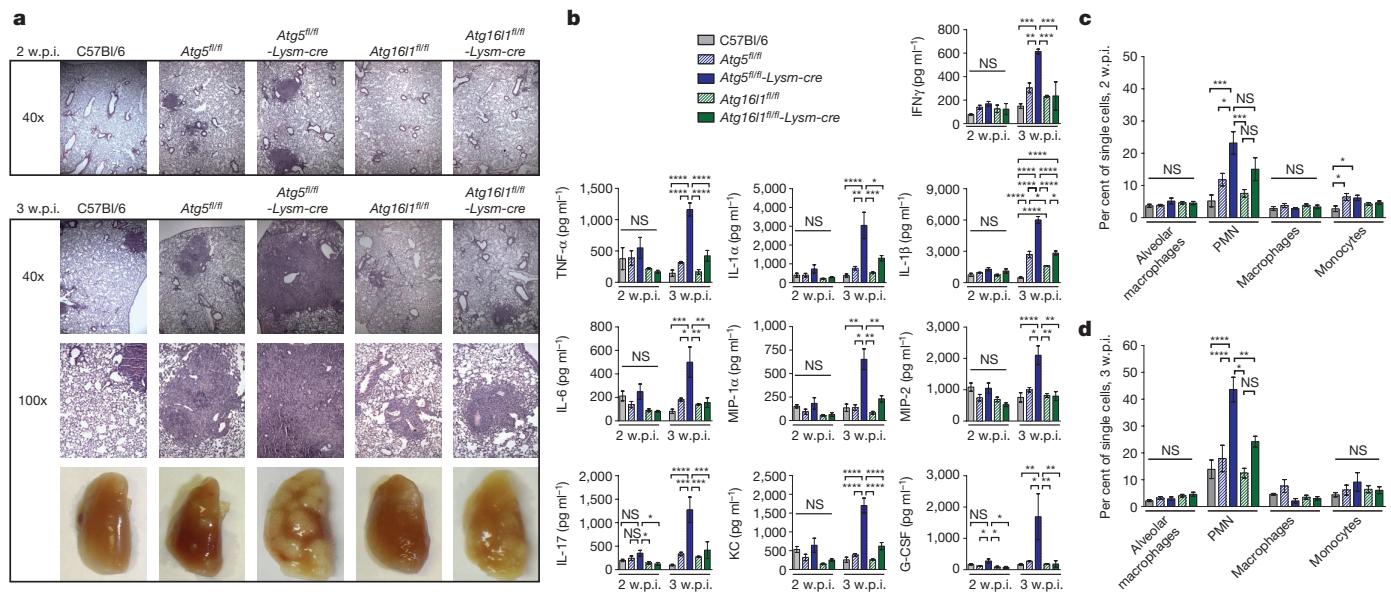


Figure 2 | Loss of *Atg5* in *LysM*⁺ cells leads to earlier and more severe lung inflammation during *M. tuberculosis* infection. **a**, Haematoxylin-and-eosin-stained histology of lungs at 2 and 3 w.p.i. and gross pathology of lungs at 3 w.p.i. **b–d**, C57Bl/6 (grey solid bars), *Atg5*^{fl/fl} (blue striped bars), *Atg5*^{fl/fl}-*Lysm-cre* (blue solid bars), *Atg16l1*^{fl/fl} (green striped bars) and *Atg16l1*^{fl/fl}-*Lysm-cre* (green solid bars). **b**, Concentration of IFN- γ , TNF- α , IL-1 α , IL-1 β , IL-6, MIP-1 α (CCL3), MIP-2 (CXCL2), IL-17a, KC (CXCL1), and G-CSF (CSF3) in lungs (homogenized in 5 ml PBS and 0.05% Tween 80) at 2 and 3 w.p.i. as detected by ELISA. **c, d**, Frequency of

alveolar macrophages, PMNs, recruited macrophages, and inflammatory monocytes as a percentage of all single cells in lungs at 2 w.p.i. (**c**) and 3 w.p.i. (**d**). Statistical differences were determined by one-way ANOVA and Bonferroni's multiple comparison test (**b–d**). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$; NS, not significant; error bars represent mean \pm s.e.m. Samples represent biological replicates. See Supplementary Fig. 3 for sample sizes and results from all statistical comparisons, Extended Data Fig. 5 for cytokine levels in uninfected lungs, and Extended Data Fig. 6 for gating strategy and number of cells in lungs.

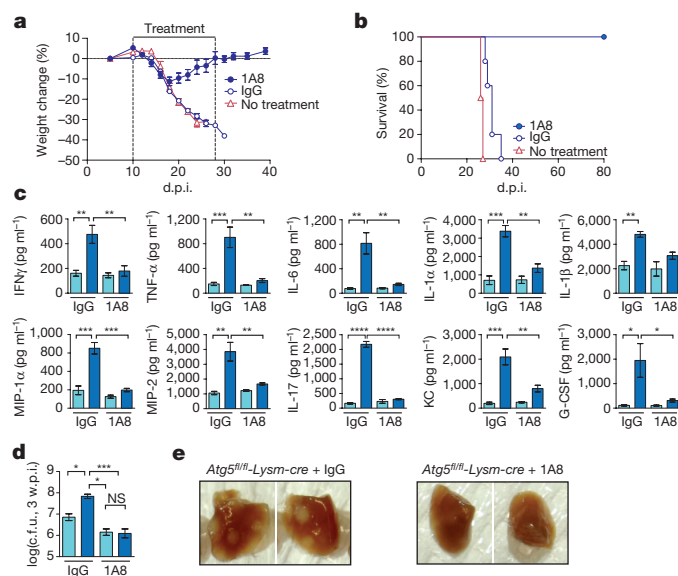


Figure 3 | Depletion of PMNs allows for survival of *Atg5*^{fl/fl}-*Lysm-cre* mice during *M. tuberculosis* infection. **a, b**, Weight change (**a**) and survival (**b**) of *Atg5*^{fl/fl}-*Lysm-cre* mice that received PMN-depleting anti-Ly6G (1A8, closed blue circle), isotype control (IgG, open blue circle), or no treatment (open pink triangle) every other day from 10–28 d.p.i. **c, d**, *Atg5*^{fl/fl} (light blue bars) and *Atg5*^{fl/fl}-*Lysm-cre* (dark blue bars) mice were treated with IgG or 1A8 and analysed at 3 w.p.i. Cytokine concentration in lungs (homogenized in 5 ml PBS plus 0.05% Tween 80) (**c**) and log pulmonary c.f.u. (**d**). **e**, Images of representative lungs from *Atg5*^{fl/fl}-*Lysm-cre* mice at 3 w.p.i. following treatment with IgG or 1A8. Statistical differences were determined by one-way ANOVA and Bonferroni's multiple comparison test (**c, d**). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$; NS, not significant; error bars represent mean \pm s.e.m. Samples represent biological replicates. See Supplementary Fig. 4 for sample sizes and results from all statistical comparisons.

Ifngr1^{-/-} mice both succumb to *M. tuberculosis* by 60 d.p.i.¹¹. Furthermore, at 3 w.p.i., PMN-depleted *Atg5*^{fl/fl}-*Lysm-cre* mice had significantly lower levels of pro-inflammatory cytokines, pulmonary burden, and lung pathology than IgG-control-treated mice (Fig. 3c–e). The depletion of PMNs alleviated all phenotypes observed at 3 w.p.i. in *Atg5*^{fl/fl}-*Lysm-cre* mice, indicating that a dysfunctional PMN response leads to the susceptibility of these mice.

We next sought to determine in which cell type(s) *Atg5* is required to control *M. tuberculosis*. *Lysm-cre* deletion occurs in PMNs, macrophages, inflammatory monocytes and myeloid-derived dendritic cells^{19,20}, indicating that *Atg5* plays a critical role in one or more of these populations during *M. tuberculosis* infection. Alveolar macrophages are the first cells infected upon inhalation of *M. tuberculosis* and are required for the establishment of infection³⁰. Furthermore, previous *in vitro* studies suggested that a predominant role for ATG5 during *M. tuberculosis* infection is to control bacterial replication in macrophages^{1–4,6–8}. Therefore, we investigated whether ATG5 is required in alveolar macrophage to control *M. tuberculosis* by infecting *Atg5*^{fl/fl}-*Cd11c-cre* mice, which lack ATG5 in alveolar macrophages and dendritic cells²⁰. In contrast to *Atg5*^{fl/fl}-*Lysm-cre* mice, *Atg5*^{fl/fl}-*Cd11c-cre* mice did not lose weight during *M. tuberculosis* infection, were able to control bacterial burden, and survived over 80 d.p.i. (Fig. 4a–c). Alveolar macrophage from *Atg5*^{fl/fl}-*Cd11c-cre* and *Atg5*^{fl/fl}-*Lysm-cre* mice displayed similar autophagy defects (Fig. 4d and Extended Data Fig. 2), indicating that resistance to *M. tuberculosis* is neither dependent on nor correlated with autophagic capacity in alveolar macrophages. Furthermore, this suggests ATG5 plays an essential role within other cells targeted by *Lysm-cre*-mediated gene deletion, such as PMNs, recruited macrophages and/or inflammatory monocytes, to control *M. tuberculosis* infection.

We have shown that excessive PMN-dominated inflammation leads to the susceptibility of *Atg5*^{fl/fl}-*Lysm-cre* mice. To determine whether loss of *Atg5* from PMN is sufficient to cause susceptibility to *M. tuberculosis*, we next used *Atg5*^{fl/fl}-*MRP8-cre* (MRP8 also known as *S100a8*) mice, which delete *Atg5* in PMNs²⁰ (Fig. 4e). *Atg5*^{fl/fl}-*MRP8-cre*

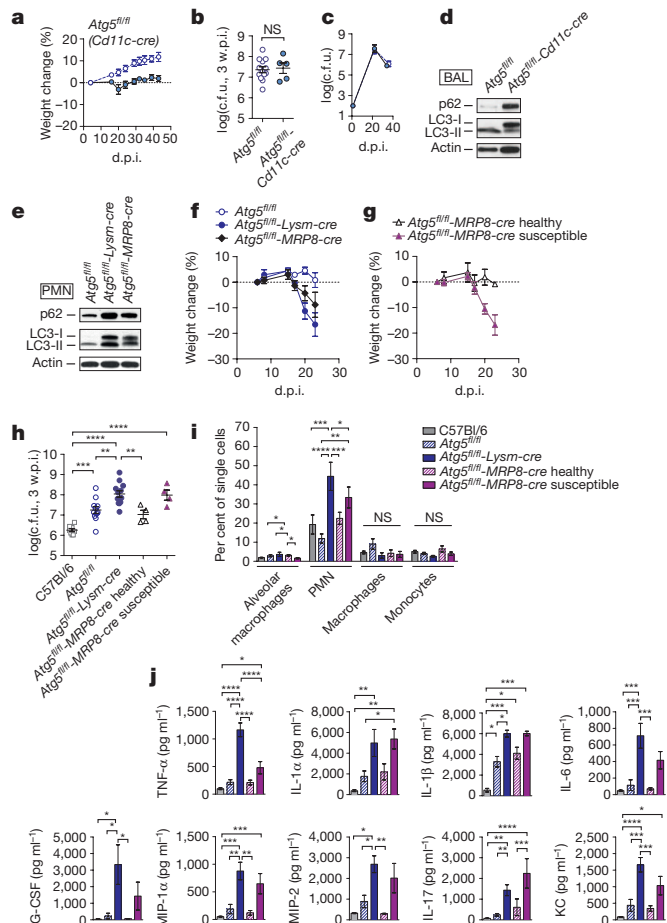


Figure 4 | Loss of *Atg5* in PMNs, but not alveolar macrophages or dendritic cells, can cause susceptibility to *M. tuberculosis*. **a–c**, Weight change (**a**) and log pulmonary c.f.u. (**b**, **c**) of *Atg5^{fl/fl}* (open circles) and *Atg5^{fl/fl}-Cd11c-cre* (closed circles). **d**, Western blot analysis of p62, LC3, and actin in bronchoalveolar macrophages (BAL) from *Atg5^{fl/fl}* and *Atg5^{fl/fl}-Cd11c-cre* mice. **e**, Western blot analysis of p62, LC3, and actin in bone marrow PMNs from *Atg5^{fl/fl}*, *Atg5^{fl/fl}-Lysm-cre* and *Atg5^{fl/fl}-MRP8-cre* mice. **f**, Weight change of *Atg5^{fl/fl}* (open blue circles), *Atg5^{fl/fl}-Lysm-cre* (closed blue circles), *Atg5^{fl/fl}-MRP8-cre* (closed black diamonds) mice following infection with *M. tuberculosis*. 50% of *Atg5^{fl/fl}-MRP8-cre* mice lost over 5% of their weight by 20 d.p.i. ('susceptible', closed purple triangles) while 50% of *Atg5^{fl/fl}-MRP8-cre* mice did not ('healthy', open black triangles). **h**, log pulmonary c.f.u. at 3 w.p.i. **i**, **j**, C57Bl/6 (grey solid bars), *Atg5^{fl/fl}* (blue striped bars), *Atg5^{fl/fl}-Lysm-cre* (blue solid bars), healthy *Atg5^{fl/fl}-MRP8-cre* (purple striped bars), and 'susceptible' *Atg5^{fl/fl}-MRP8-cre* (purple solid bars). **i**, Concentration of TNF- α , IL-1 α , IL-1 β , IL-6, MIP-1 α (CCL3), MIP-2 (CXCL2), IL-17 α , KC (CXCL1), and G-CSF (CSF3) in lungs (homogenized in 5 ml PBS + 0.05% Tween 80) at 3 w.p.i. **j**, Frequency of alveolar macrophages, PMNs, recruited macrophages, and inflammatory monocytes as a percentage of single cells in lungs at 3 w.p.i. Statistical differences were determined by one-way ANOVA and Bonferroni's multiple comparison test (**h–j**). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$; NS, not significant; error bars represent mean \pm s.e.m. Samples represent biological replicates. See Supplementary Fig. 5 for sample sizes and results from all statistical comparisons, and Extended Data Fig. 7 for total numbers of cells in lungs.

mice were more susceptible to *M. tuberculosis* infection, as indicated by an average increase in weight loss compared to *Atg5^{fl/fl}* controls (Fig. 4f). However, analysis of individual mice revealed that only half of the *Atg5^{fl/fl}-MRP8-cre* mice lost weight following *M. tuberculosis* infection (between 10–20% of their starting weight); the remaining mice exhibited an average 2% weight gain. This split phenotype was reproducible across multiple experiments, and was independent of differences

in age, sex or litter of the mice, suggesting a threshold effect in the susceptibility of the *Atg5^{fl/fl}-MRP8-cre* mice. To study these two distinct outcomes, we compared responses in mice that lost over 5% of their starting weight at 20 d.p.i. ('susceptible') with the remaining mice ('healthy') (Fig. 4g). At 3 w.p.i., lungs from susceptible *Atg5^{fl/fl}-MRP8-cre* mice exhibited higher bacterial burden, cytokine responses, and frequency of PMNs (Fig. 4h–j and Extended Data Fig. 7). The susceptible *Atg5^{fl/fl}-MRP8-cre* mice displayed the same phenotypes as the *Atg5^{fl/fl}-Lysm-cre* mice, demonstrating a PMN-intrinsic role for ATG5 during acute *M. tuberculosis* infection. However, the incomplete penetrance of susceptibility in *Atg5^{fl/fl}-MRP8-cre* mice suggests that the extreme sensitivity of *Atg5^{fl/fl}-Lysm-cre* mice to *M. tuberculosis* results from the loss of ATG5 in macrophage and monocytes, as well as PMNs. Notably, *Atg16l1^{fl/fl}-Lysm-cre* mice are not susceptible to *M. tuberculosis* infection, even though PMNs (Extended Data Fig. 8), in addition to macrophages (Fig. 1l and Extended Data Fig. 2), from *Atg16l1^{fl/fl}-Lysm-cre* and *Atg5^{fl/fl}-Lysm-cre* mice have a similar defect in autophagy. This further supports that ATG5 functions, at least in part, independently of ATG16L1 to protect mice from *M. tuberculosis* infection.

Despite numerous *in vitro* studies emphasizing a role for autophagy in macrophages during *M. tuberculosis* infection (including, but not limited to, refs 1–8), our data show that loss of genes essential for canonical autophagy does not correlate with susceptibility to *M. tuberculosis* in the context of a complete immune response in the host. Importantly, mice used in our studies have similar autophagy defects and have been used in prior publications to investigate the function of individual ATG factors^{12–15}, validating these mice as suitable genetic models to study autophagy *in vivo*. Our studies indicate that prior reports analysing the role of only a single autophagy gene to conclude that canonical autophagy is responsible for the phenotypes observed need to be re-examined. The observation that the *Atg5^{fl/fl}* and *Atg5^{fl/fl}-Lysm-cre* mice have only small differences in *M. tuberculosis* burden supports the other data presented here that the dramatic difference in the inflammatory response is the predominant driver of susceptibility in *Atg5^{fl/fl}-Lysm-cre* mice during *M. tuberculosis* infection. The apparent insignificance of autophagy for controlling *M. tuberculosis* replication may reflect the fact that *M. tuberculosis* encodes highly effective inhibitors of canonical autophagy; however, these mechanisms have yet to be described. Furthermore, studies investigating loss of autophagy, including this one, do not address whether activation of autophagy could enhance restriction of *M. tuberculosis* replication.

By analysing different Cre-mediated deletion strains, we have found that loss of ATG5 in PMNs, but not alveolar macrophages or dendritic cells, can result in the loss of control of *M. tuberculosis* infection, but the severe susceptibility of the *Atg5^{fl/fl}-Lysm-cre* mice relies on deletion of ATG5 in multiple LysM⁺ cell types. These data also reveal a PMN-intrinsic role for ATG5 during *M. tuberculosis* infection. Importantly, the reversal of all phenotypes in the *Atg5^{fl/fl}-Lysm-cre* mice upon PMN depletion positions PMN as a major driver in the dysfunctional response in these mice. Our experiments suggest a model where infection with *M. tuberculosis* induces a pro-inflammatory response that leads to the recruitment of PMNs to the lung. The absence of ATG5 expression within the responding myeloid cells leads to uncontrolled accumulation of PMNs in the lung, which causes increased pathology and probably provides an expanded niche for bacterial infection. The animal then succumbs to infection before the adaptive immune response is able to control the inflammation and bacterial replication. Together, the *in vivo* genetic analyses presented here argue for a shift in focus onto macroautophagy-independent roles of ATG5 in controlling resistance to *M. tuberculosis* infection *in vivo*.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 February; accepted 16 November 2015.

Published online 9 December 2015.

1. Deretic, V. Autophagy in tuberculosis. *Cold Spring Harb. Persp. Med.* **4**, a018481 (2014).
2. Dutta, R. K., Kathania, M., Raj, M. & Majumdar, S. IL-6 inhibits IFN- γ induced autophagy in *Mycobacterium tuberculosis* H37Rv infected macrophages. *Int. J. Biochem. Cell Biol.* **44**, 942–954 (2012).
3. Juárez, E. *et al.* NOD2 enhances the innate response of alveolar macrophages to *Mycobacterium tuberculosis* in humans. *Eur. J. Immunol.* **42**, 880–889 (2012).
4. Watson, R. O., Manzanillo, P. S. & Cox, J. S. Extracellular *M. tuberculosis* DNA targets bacteria for autophagy by activating the host DNA-sensing pathway. *Cell* **150**, 803–815 (2012).
5. Seto, S., Tsujimura, K., Horii, T. & Koide, Y. Autophagy adaptor protein p62/SQSTM1 and autophagy-related gene Atg5 mediate autophagosome formation in response to *Mycobacterium tuberculosis* infection in dendritic cells. *PLoS ONE* **8**, e86017 (2013).
6. Sakowski, E. T. *et al.* Ubiquitin 1 promotes IFN- γ -induced xenophagy of *Mycobacterium tuberculosis*. *PLoS Pathog.* **11**, e1005076 (2015).
7. Gutierrez, M. G. *et al.* Autophagy is a defense mechanism inhibiting BCG and *Mycobacterium tuberculosis* survival in infected macrophages. *Cell* **119**, 753–766 (2004).
8. Wang, J. *et al.* MicroRNA-155 promotes autophagy to eliminate intracellular mycobacteria by targeting Rheb. *PLoS Pathog.* **9**, e1003697 (2013).
9. Castillo, E. F. *et al.* Autophagy protects against active tuberculosis by suppressing bacterial burden and inflammation. *Proc. Natl Acad. Sci. USA* **109**, E3168–E3176 (2012).
10. Cooper, A. M. *et al.* Disseminated tuberculosis in interferon gamma gene-disrupted mice. *J. Exp. Med.* **178**, 2243–2247 (1993).
11. Nandi, B. & Behar, S. M. Regulation of neutrophils by interferon- γ limits lung inflammation during tuberculosis infection. *J. Exp. Med.* **208**, 2251–2262 (2011).
12. Choi, J. *et al.* The parasitophorous vacuole membrane of *Toxoplasma gondii* is targeted for disruption by ubiquitin-like conjugation systems of autophagy. *Immunity* **40**, 924–935 (2014).
13. Hwang, S. *et al.* Nondegradative role of Atg5-Atg12/Atg16L1 autophagy protein complex in antiviral activity of interferon gamma. *Cell Host Microbe* **11**, 397–409 (2012).
14. Zhao, Z. *et al.* Autophagosome-independent essential function for the autophagy protein Atg5 in cellular immunity to intracellular pathogens. *Cell Host Microbe* **4**, 458–469 (2008).
15. Martinez, J. *et al.* Molecular characterization of LC3-associated phagocytosis reveals distinct roles for Rubicon, NOX2 and autophagy proteins. *Nature Cell Biol.* **17**, 893–906 (2015).
16. Jounai, N. *et al.* The Atg5 Atg12 conjugate associates with innate antiviral immune responses. *Proc. Natl Acad. Sci. USA* **104**, 14050–14055 (2007).
17. Yousefi, S. *et al.* Calpain-mediated cleavage of Atg5 switches autophagy to apoptosis. *Nature Cell Biol.* **8**, 1124–1132 (2006).
18. Maskay, D. *et al.* ATG5 is induced by DNA-damaging agents and promotes mitotic catastrophe independent of autophagy. *Nature Commun.* **4**, 2130 (2013).
19. Jakubzick, C. *et al.* Lymph-migrating, tissue-derived dendritic cells are minor constituents within steady-state lymph nodes. *J. Exp. Med.* **205**, 2839–2850 (2008).
20. Abram, C. L., Roberge, G. L., Hu, Y. & Lowell, C. A. Comparative analysis of the efficiency and specificity of myeloid-Cre deleting strains using ROSA-EYFP reporter mice. *J. Immunol. Methods* **408**, 89–100 (2014).
21. Flynn, J. L. & Chan, J. Immunology of tuberculosis. *Annu. Rev. Immunol.* **19**, 93–129 (2001).
22. Jayaswal, S. *et al.* Identification of host-dependent survival factors for intracellular *Mycobacterium tuberculosis* through an siRNA Screen. *PLoS Pathog.* **6**, e1000839 (2010).
23. Jung, C. H. *et al.* ULK-Atg13-FIP200 complexes mediate mTOR signaling to the autophagy machinery. *Mol. Biol. Cell* **20**, 1992–2003 (2009).
24. Mariño, G. *et al.* Autophagy is essential for mouse sense of balance. *J. Clin. Invest.* **120**, 2331–2344 (2010).
25. Lin, H. H. *et al.* Dynamic involvement of ATG5 in cellular stress responses. *Cell Death Dis.* **5**, e1478 (2014).
26. Cadwell, K. *et al.* A key role for autophagy and the autophagy gene *Atg16l1* in mouse and human intestinal Paneth cells. *Nature* **456**, 259–263 (2008).
27. Abdel Fattah, E., Bhattacharya, A., Herron, A., Safdar, Z. & Eissa, N. T. Critical role for IL-18 in spontaneous lung inflammation caused by autophagy deficiency. *J. Immunol.* **194**, 5407–5416 (2015).
28. Kanayama, M., He, Y.-W. & Shinohara, M. L. The lung is protected from spontaneous inflammation by autophagy in myeloid cells. *J. Immunol.* **194**, 5465–5471 (2015).
29. Saitoh, T. *et al.* Loss of the autophagy protein Atg16L1 enhances endotoxin-induced IL-1 β production. *Nature* **456**, 264–268 (2008).
30. Leemans, J. C. *et al.* Depletion of alveolar macrophages exerts protective effects in pulmonary tuberculosis in mice. *J. Immunol.* **166**, 4604–4611 (2001).

Supplementary Information is available in the online version of the paper.

Acknowledgements C.L.S. is supported by a Beckman Young Investigator Award from the Arnold and Mabel Beckman Foundation. J.M.K. is supported by a National Science Foundation Graduate Research Fellowship DGE-1143954 and the NIGMS Cell and Molecular Biology Training Grant GM007067. J.P.H. is supported by a National Science Foundation Graduate Research Fellowship DGE-1143954. H.W.V., S.P. and A.K. are supported by U19 AI109725. We would like to acknowledge D. Kreamealmeyer for assistance with the mouse colonies and T. Malek and L. D. Sibley for providing the 1A8 hybridoma.

Author Contributions J.M.K. designed and performed experiments, analysed data, and wrote the manuscript. J.P.H. and L.A.W. performed experiments. S.P., A.K. and J.D. generated mouse strains. H.W.V. provided all mouse strains, analysed data and wrote the manuscript. C.L.S. designed experiments, analysed data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.L.S. (stallings@wustm.wustl.edu).

METHODS

No statistical methods were used to predetermine sample size.

Cells and media. *Mycobacterium tuberculosis* Erdman was cultured at 37°C in 7H9 (broth) or 7H10 (agar) (Difco) medium supplemented with 10% oleic acid/albumin/dextrose/catalase (OADC), 0.5% glycerol, and 0.05% Tween 80 (broth).

Ex vivo macrophages were enriched from mice via bronchoalveolar lavage or peritoneal lavage with DMEM + 10% FBS + 1% MEM non-essential amino acids (Cellgro 25-025-CI) + 100 U ml⁻¹ penicillin + 100 mg ml⁻¹ streptomycin (Sigma P4333). Lavage cells were treated with ACK lysis buffer (0.15 M NH₄Cl, 10 mM KHCO₃, 0.1 mM EDTA) to lyse red blood cells, plated in tissue-culture-treated plates, and incubated at 37°C in 5% CO₂ for at least 4 h to allow adherence of macrophages¹⁴. Wells were washed vigorously with PBS to remove non-adherent cells and lysed in 2× Laemmli buffer for western blot analysis.

Bone-marrow-derived macrophages were isolated from femurs and tibias of mice, and cultured in DMEM + 20% FBS + 10% supernatant from 3T3 cells overexpressing M-CSF + 1% MEM non-essential amino acids (Cellgro 25-025-CI) + 100 U ml⁻¹ penicillin and 100 µg ml⁻¹ streptomycin (Sigma P4333) at 37°C in 5% CO₂.

PMNs for *ex vivo* western blotting analysis were purified from uninfected bone marrow by negative selection via MACS column (Miltenyi Biotec, 130-097-658) according to the manufacturer's guidelines and immediately lysed in 2× Laemmli buffer.

Western blotting. Protein samples were diluted in 2× Laemmli buffer, resolved using 4–20% polyacrylamide gels (BioRad no. 456-1096) transferred to PVDF membrane (GE Healthcare 10600023) and detected with the following antibodies: LC3b (Sigma L7543—detects LC3-I and LC3-II), p62/SQSTM1 (Sigma P0067), ATG5 (Sigma A2859), β-actin (Cell Signaling Technology no. 4970) and goat-anti-mouse-horseradish peroxidase (HRP) and goat-anti-rabbit-HRP as appropriate. HRP was detected using Western Lightning Plus ECL (PerkinElmer no. NEL103001EA) for actin or ECL Prime (GE Healthcare RPN2232) for LC3b, p62 and ATG5. For gel source data, see Supplementary Fig. 1.

Mouse strains. Adult mice (7–15 weeks of age) of both sexes were used and mouse experiments were randomized. No blinding was performed during animal experiments. All mice used have been fully backcrossed to a C57Bl/6 background. Sample sizes are detailed in Supplementary Figs 2–10 and were sufficient to detect differences as small as 10% using the statistical methods described. *Atg5^{fl/fl}-Lysm-cre*, *Atg7^{fl/fl}-Lysm-cre*, *Atg16l1^{fl/fl}-Lysm-cre*, *Atg3^{fl/fl}-Lysm-cre*, and *Atg14l1^{fl/fl}-Lysm-cre*, and all floxed control mice have been previously described^{12–14}. *Atg14l1^{fl/fl}-Lysm-cre* and floxed control mice were provided by S. Akira (Osaka University, Japan). *Atg3^{fl/fl}-Lysm-cre* mice were derived from *Atg3^{fl/fl}* mice provided by Y.-W. He (Duke University, USA)³¹. *Atg12^{fl/fl}-Lysm-cre* mice were derived from *Atg12^{fl/fl}* mice³². *Atg16l1^{HM1}* mice have been previously described (HM1, BC0122 strain)²⁶. *p62^{-/-}* mice were supplied by E. White (Rutgers University, USA)³³. *Atg4b^{g/gt}* mice were previously described, and are referred to as *Atg4b^{-/-}* throughout the text²⁴. *Ulk1^{-/-}* and *Ulk2^{-/-}* mice were provided by S. Tooze (London Research Institute, UK)^{34,35}. *Atg5^{fl/fl}-Cd11c-cre* and *Atg5^{fl/fl}-MRP8-cre* mice were generated in our facility by crossing *Atg5^{fl/fl}* to *Cd11c-cre* (The Jackson Laboratory 007567) and *MRP8-cre* (The Jackson Laboratory 021614).

All procedures involving animals were conducted following the National Institutes of Health guidelines for housing and care of laboratory animals and performed in accordance with institutional regulations after protocol review and approval by the Institutional Animal Care and Use Committee of The Washington University in St. Louis School of Medicine (protocol no. 20130156, Analysis of Mycobacterial Pathogenesis). Washington University is registered as a research facility with the United States Department of Agriculture and is fully accredited by the American Association of Accreditation of Laboratory Animal Care. The Animal Welfare Assurance is on file with OPRR-NIH. All animals used in these experiments were subjected to no or minimal discomfort. All mice were euthanized by CO₂ asphyxiation, which is approved by the American Veterinary Association Panel on Euthanasia.

***M. tuberculosis* infection of mice.** Before infection, exponentially replicating *M. tuberculosis* were washed in PBS + 0.05% Tween 80, and sonicated to disperse clumps. 7- to 15-week-old female and male mice were exposed to 8 × 10⁷ colony-forming units (c.f.u.) of *M. tuberculosis* in an Inhalation Exposure System (Glas-Col), which delivers ~100 bacteria to the lung per animal. At 24 h after infection, the bacterial titres in the lungs of at least two mice were determined to confirm the dose of *M. tuberculosis* inoculation. The dose determined from

these mice is assumed to represent the average dose received by all mice in the same infection. Bacterial burden was determined by plating serial dilutions of lung homogenates onto 7H10 agar plates. Plates were incubated at 37°C in 5% CO₂ for 3 weeks before counting colonies.

Flow cytometry. Lungs were perfused with sterile PBS and digested at 37°C for 1 h with 625 µg ml⁻¹ collagenase D (Roche 11088875103) and 75 U ml⁻¹ DNase I (Sigma D4527). Single-cell suspensions were stained in PBS + 2% FBS + 0.1% sodium azide in the presence of Fc receptor blocking antibody (BD Pharmingen 553541) and stained with the antibodies against the following mouse markers: CD11b_PerCP-Cy5.5 (BD Pharmingen 550993), CD11c_APC-Cy7 (eBioscience 47-0114), Ly6C_PE (BD Pharmingen 560592), Ly6G_PE-Cy7 (BD Pharmingen 560601), and F4/80_APC (Invitrogen MF48005). The FITC channel was used to determine autofluorescence. Cells were stained for 20 min at 4°C and then fixed in 4% paraformaldehyde (Electron Microscopy Sciences) for 20 min at room temperature. Flow cytometry was performed on a FACSCanto II (BD Bioscience) and data was analysed with FlowJo (Tree Star Inc.). Gating strategies are depicted in Extended Data Fig. 6a.

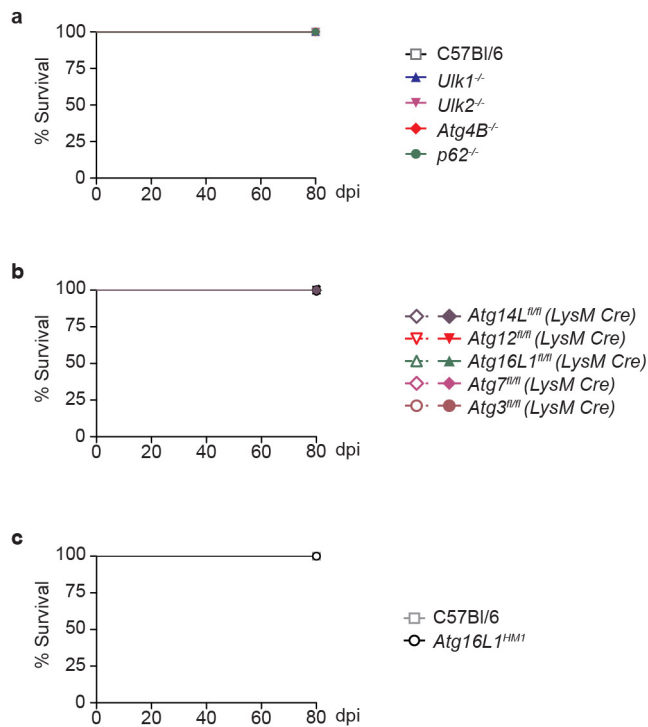
Cytokine analysis. Lungs (right lobe) were homogenized in 1 ml (uninfected mice) or 5 ml (*M. tuberculosis*-infected mice) PBS + 0.05% Tween 80. Homogenized tissue supernatants were filtered (0.22 µm) and analysed by ELISA according to the manufacturer's guidelines (R&D systems): KC/CXCL1 (DY453), IFN-γ (DY485), TNF-α (DY410), IL-1α (DY400), IL-1β (DY401), IL-6 (DY406), IL-17a (DY421), MIP-1α/CCL3 (DY450), MIP-2/CXCL2 (DY452), and G-CSF (DY414).

RNA extraction and quantification. Following tissue disruption by bead-beating (MP Biosystems), RNA was extracted from *M. tuberculosis*-infected lungs using the RNeasy Kit according to the manufacturer's guidelines (Qiagen 74106). cDNA was made with SuperScript III reverse transcriptase using oligo-dT primers (Life Technologies 18080-051). qRT-PCR was performed using iTAQ SYBR Green (BioRad 172-5121) and transcript levels were normalized to actin. The following primers were used: *Atg16l1* forward, 5'-CCGAATCTGGACTGTGGATG-3'; reverse, 5'-CGGAGATCCCAGAGTTTGTAG-3'; actin forward, 5'-ACCTTCTACAATGAGCTGCG-3'; reverse, 5'-CTGGATGGCTACGTACATGG-3'.

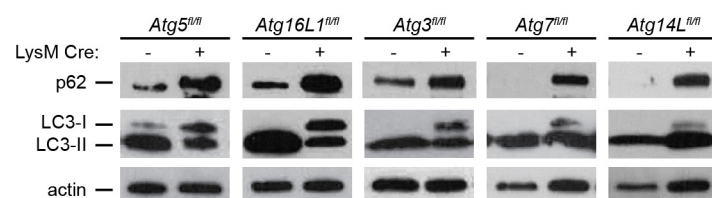
PMN depletion. Mice were treated with 0.2 mg anti-Ly6G (clone 1A8) or 0.2 mg rat IgG (Sigma I8015) via intraperitoneal injection every 48 h between days 10 and 28 post infection. Efficacy of PMN depletion was confirmed by loss of CD11b⁺ Gr-1^{high} cells in lungs at 21 d.p.i. Anti-Ly6G was collected from 1A8 hybridoma³⁶ grown in Serum Free Medium (Gibco no. 12045-076) in CL350 Bioreactor flasks (Argos Technologies no. 900 10).

Data and statistics. All experiments were performed at least twice. When shown, multiple samples represent biological (not technical) replicates of mice randomly sorted into each experimental group. No blinding was performed during animal experiments. Animals were only excluded when pathology unrelated to *M. tuberculosis* infection was present (that is, weight loss due to malocclusion). Determination of statistical differences was performed with Prism 5 (Graphpad Software, Inc.) using log-rank Mantel-Cox tests (survival), unpaired two-tailed *t*-tests (to compare two groups with similar variances), or one-way ANOVA with Bonferroni's multiple comparison test (to compare more than two groups). Sample sizes were sufficient to detect differences as small as 10% using the statistical methods described. When used, centre values and error bars represent the mean ± s.e.m. Sample sizes and the results of all comparisons can be found in Supplementary Figs 2–10.

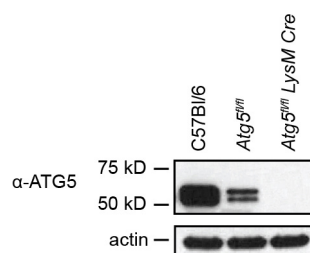
- Jia, W. & He, Y.-W. Temporal regulation of intracellular organelle homeostasis in T lymphocytes by autophagy. *J. Immunol.* **186**, 5313–5322 (2011).
- Malhotra, R., Warne, J. P., Salas, E., Xu, A. W. & Debnath, J. Loss of Atg12, but not Atg5, in pro-opiomelanocortin neurons exacerbates diet-induced obesity. *Autophagy* **11**, 145–154 (2015).
- Komatsu, M. *et al.* Homeostatic levels of p62 control cytoplasmic inclusion body formation in autophagy-deficient mice. *Cell* **131**, 1149–1163 (2007).
- Lee, E.-J. & Tournier, C. The requirement of uncoordinated 51-like kinase 1 (ULK1) and ULK2 in the regulation of autophagy. *Autophagy* **7**, 689–695 (2011).
- McAlpine, F., Williamson, L. E., Tooze, S. A. & Chan, E. Y. W. Regulation of nutrient-sensitive autophagy by uncoordinated 51-like kinases 1 and 2. *Autophagy* **9**, 361–373 (2013).
- Fleming, T. J., Fleming, M. L. & Malek, T. R. Selective expression of Ly-6G on myeloid lineage cells in mouse bone marrow. RB6-8C5 mAb to granulocyte-differentiation antigen (Gr-1) detects members of the Ly-6 family. *J. Immunol.* **151**, 2399–2408 (1993).



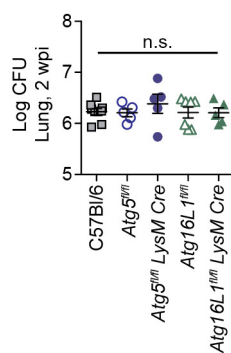
Extended Data Figure 1 | Survival of mice with defects in autophagy genes other than *Atg5*. Per cent survival of mice following infection with 100 colony-forming units (c.f.u.) of aerosolized *M. tuberculosis*. **a**, Survival of C57Bl/6 (open squares), *Ulk1*^{-/-} (blue triangles), *Ulk2*^{-/-} (inverted pink triangles), *Atg4b*^{-/-} (red diamonds), and *p62*^{-/-} (green circles) mice. **b**, Survival of *Atg14L*^{fl/fl}-*Lysm-cre* (purple diamonds), *Atg12*^{fl/fl}-*Lysm-cre* (red inverted triangles), *Atg16L1*^{fl/fl}-*Lysm-cre* (green triangles), *Atg7*^{fl/fl}-*Lysm-cre* (pink diamonds), *Atg3*^{fl/fl}-*Lysm-cre* (brown circles) and corresponding floxed control mice. Floxed control mice are shown in open shapes, LysM-Cre-expressing mice are shown in closed shapes. **c**, Survival of C57Bl/6 (open squares), *Atg16L1*^{HM1} (open circles). Samples represent biological replicates. See Supplementary Fig. 6 for sample sizes.



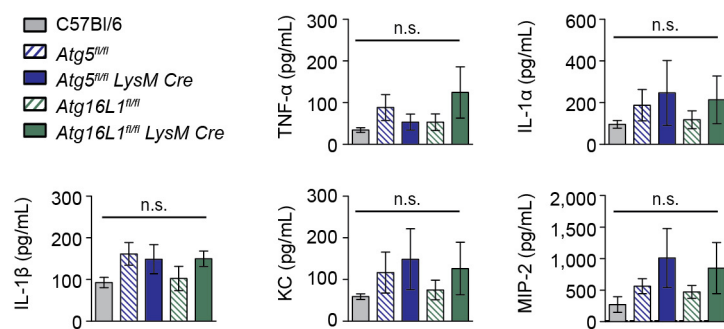
Extended Data Figure 2 | Analysis of autophagy in bronchoalveolar macrophages. Western blot analysis of p62, LC3, and actin levels in *ex vivo* macrophages isolated from bronchoalveolar lavages of uninfected mice. For gel source data, see Supplementary Fig. 1.



Extended Data Figure 3 | *Atg5^{fl/fl}* bone-marrow-derived macrophages are hypomorphic for ATG5. Western blot analysis of ATG5 (ATG5–ATG12 conjugate, 56 kDa) and actin in uninfected-bone-marrow-derived macrophages. For gel source data, see Supplementary Fig. 1.



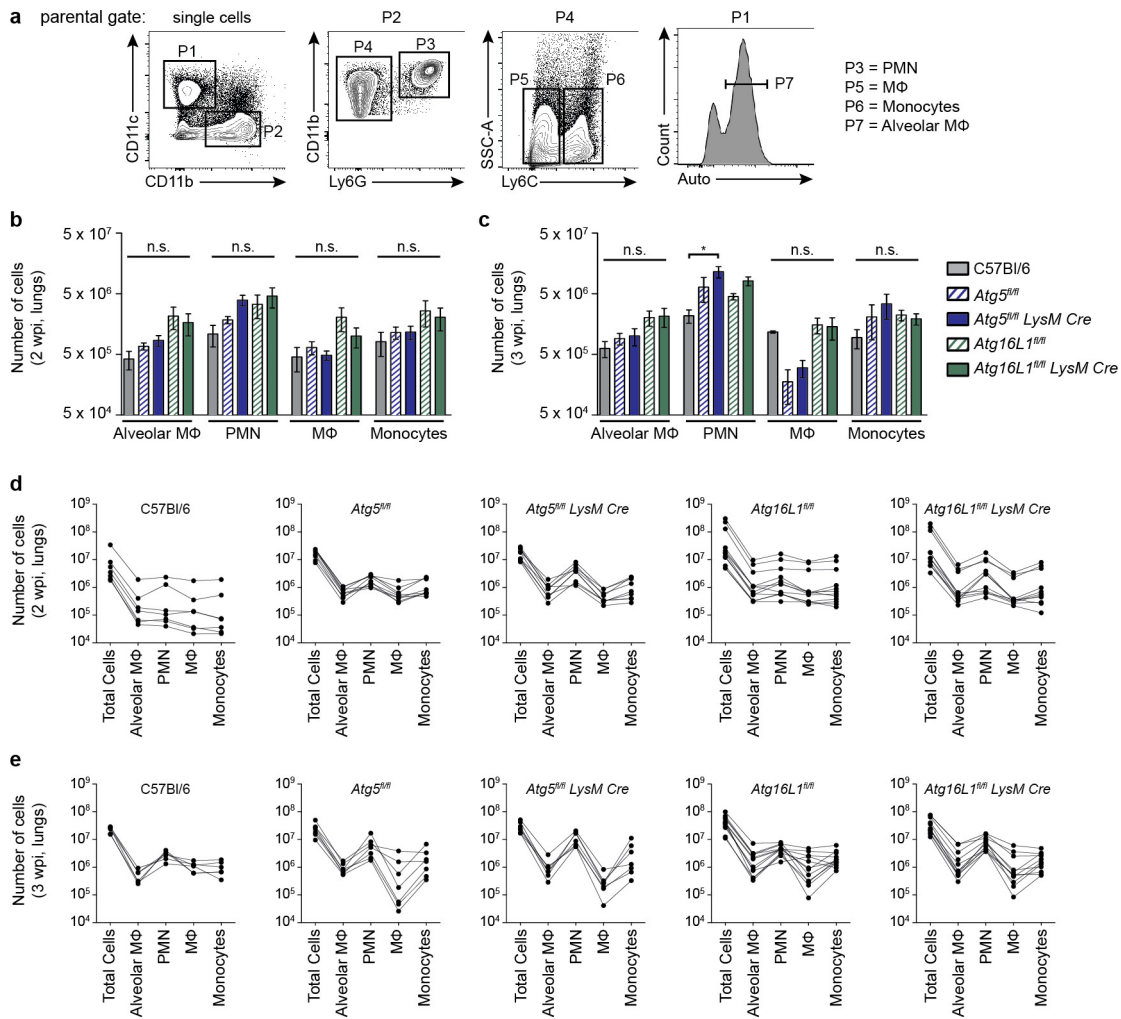
Extended Data Figure 4 | Loss of *Atg5* or *Atg16L1* in *LysM*⁺ cells does not lead to increased c.f.u. at 2 w.p.i. log pulmonary c.f.u. at 2 w.p.i.
 Samples represent biological replicates; error bars represent mean \pm s.e.m.
 See Supplementary Fig. 7 for sample sizes and results from all statistical comparisons.



Extended Data Figure 5 | Cytokine levels in uninfected lungs.

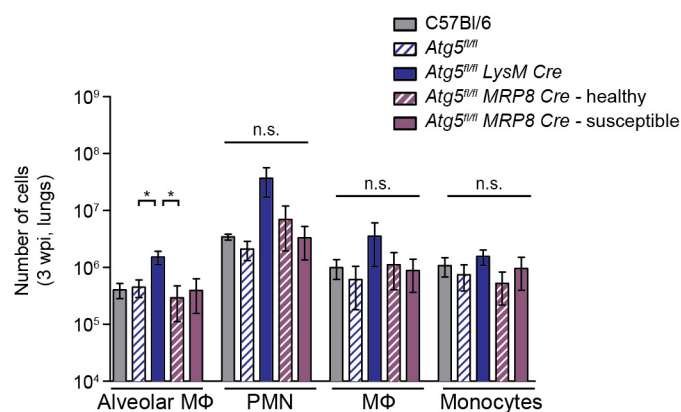
Concentration of cytokines in lungs (homogenized in 1 ml PBS plus 0.05% Tween 80) from uninfected mice. Levels of IFN γ , IL-6, MIP-1 α , IL-17, and G-CSF were below the limit of detection. C57Bl/6 (grey solid bars), *Atg5^{fl/fl}* (blue striped bars), *Atg5^{fl/fl}-LysM-cre* (blue solid bars), *Atg16L1^{fl/fl}*

(green striped bars), *Atg16L1^{fl/fl}-LysM-cre* (green solid bars). Statistical differences were determined by one-way ANOVA and Bonferonni's multiple comparison test. n.s., not significant. Samples represent biological replicates; error bars represent mean \pm s.e.m. See Supplementary Fig. 8 for sample sizes and results from all statistical comparisons.

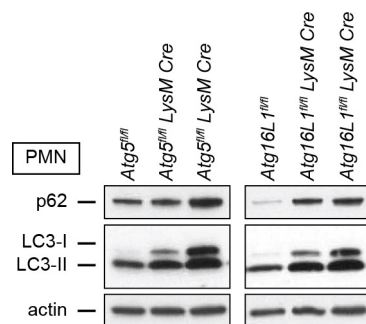


Extended Data Figure 6 | Number of inflammatory cells in lungs of mice at 2 and 3 w.p.i. (related to Fig. 2). **a**, Gating strategy for analysis of inflammatory cells in lungs at 2 and 3 w.p.i. Single lung cells were gated based on CD11b, CD11c, Ly6G, Ly6C and autofluorescence (auto). The parental gate is shown above each contour plot. Representative data are shown from an *Atg5^{fl/fl}* mouse at 2 w.p.i. **b**, **c**, C57Bl/6 (grey solid bars), *Atg5^{fl/fl}* (blue striped bars), *Atg5^{fl/fl}-LysM-cre* (blue solid bars), *Atg16L1^{fl/fl}* (green striped bars), *Atg16L1^{fl/fl}-LysM-cre* (green solid bars). Mean number of alveolar macrophages, PMNs, recruited macrophages, and inflammatory monocytes in lungs at 2 w.p.i. (**b**) and 3 w.p.i. (**c**). **d**, **e**, Flow cytometry data presented in **b** and **c** and in Fig. 2 are the compilation of results from five experiments. In some experiments, different amounts of

lung were collected for analysis, making it difficult to compare the average number of each cell type between strains, unless the data are normalized (as done in Fig. 2c, d—percentage of total cells). Therefore, to compare the raw number of cells detected in each cell population, each mouse analysed at 2 w.p.i. (**d**) and 3 w.p.i. (**e**) has been graphed individually. Each line represents a different mouse, with dots indicating the number of total cells, alveolar macrophages, PMNs, recruited macrophages and inflammatory monocytes. Statistical differences were determined by one-way ANOVA and Bonferonni's multiple comparison test (**b**, **c**); * $P < 0.05$; n.s., not significant. Samples represent biological replicates; error bars represent mean \pm s.e.m. See Supplementary Fig. 9 for sample sizes and results from all statistical comparisons.



Extended Data Figure 7 | Number of inflammatory cells in lungs of mice at 3 w.p.i. (related to Fig. 4). Number of alveolar macrophages, PMNs, recruited macrophages, and inflammatory monocytes in lungs at 3 w.p.i. C57Bl/6 (grey solid bars), *Atg5^{fl/fl}* (blue striped bars), *Atg5^{fl/fl} LysM Cre* (blue solid bars), 'healthy' *Atg5^{fl/fl} MRP8 Cre* (purple striped bars), and 'susceptible' *Atg5^{fl/fl} MRP8 Cre* (purple solid bars). Statistical differences were determined by one-way ANOVA and Bonferroni's multiple comparison test; * $P < 0.05$; n.s., not significant. Samples represent biological replicates; error bars represent mean \pm s.e.m. See Supplementary Fig. 10 for sample sizes and results from all statistical comparisons.



Extended Data Figure 8 | Analysis of autophagy in bone marrow PMNs.

Western blot analysis of p62, LC3, and actin in bone marrow PMNs from uninfected mice. Each lane represents an individual mouse. Two replicates of the *Atg5^{fl/fl}-Lysm-cre* and *Atg16L1^{fl/fl}-Lysm-cre* mice are shown. For gel source data, see Supplementary Fig. 1.

Germline variant FGFR4 p.G388R exposes a membrane-proximal STAT3 binding site

Vijay K. Ulaganathan¹, Bianca Sperl¹, Ulf R. Rapp² & Axel Ullrich¹

Variant rs351855-G/A is a commonly occurring single-nucleotide polymorphism of coding regions in exon 9 of the fibroblast growth factor receptor *FGFR4* (*CD334*) gene (c.1162G>A). It results in an amino-acid change at codon 388 from glycine to arginine (p.Gly388Arg) in the transmembrane domain of the receptor. Despite compelling genetic evidence for the association of this common variant with cancers of the bone¹, breast², colon³, prostate^{4,5}, skin⁶, lung^{7,8}, head and neck⁹, as well as soft-tissue sarcomas and non-Hodgkin lymphoma, the underlying biological mechanism has remained elusive. Here we show that substitution of the conserved glycine 388 residue to a charged arginine residue alters the transmembrane spanning segment and exposes a membrane-proximal cytoplasmic signal transducer and activator of transcription 3 (STAT3) binding site Y³⁹⁰-(P)XXQ³⁹³. We demonstrate that such membrane-proximal STAT3 binding motifs in the germline of type I membrane receptors enhance STAT3 tyrosine phosphorylation by recruiting STAT3 proteins to the inner cell membrane. Remarkably, such germline variants frequently co-localize with somatic mutations in the Catalogue of Somatic Mutations in Cancer (COSMIC) database. Using *Fgfr4* single nucleotide polymorphism knock-in mice and transgenic mouse models for breast and lung cancers, we validate the enhanced STAT3 signalling induced by the FGFR4 Arg388-variant *in vivo*. Thus, our findings elucidate the molecular mechanism behind the genetic association of rs351855 with accelerated cancer progression and suggest that germline variants of cell-surface molecules that recruit STAT3 to the inner cell membrane are a significant risk for cancer prognosis and disease progression.

We previously reported that the FGFR4 Arg388 allele (rs351855-A) located in exon 9 of the *FGFR4* gene (Extended Data Fig. 1a) is associated with cancer progression and poor prognosis^{6,7,9,10}. According to the current data set from the 1000 Genomes Project¹¹, this common variant of FGFR4 occurs at a global minor allele frequency of 0.30 (Extended Data Fig. 1b), and its frequency is approximately 50% in patients with cancer¹⁰. To identify the pathogenic signalling events specific to the *FGFR4* p.G388R variant, we generated C57BL/6 knock-in mice harbouring the mouse homologue of the *FGFR4* p.G388R allele corresponding to the human single nucleotide polymorphism (SNP) rs351855-G/A (hereafter, the genotype is denoted as *Fgfr4*^{G/G} for wild type and *Fgfr4*^{A/A} for risk variant, and the encoded mouse homologue protein is denoted as FGFR4 Gly385 and FGFR4 Arg385 respectively). The *Fgfr4* SNP knock-in mouse was verified by nucleotide sequencing (Extended Data Fig. 1c). The expression levels of the *Fgfr4* mutant transcript and protein were identical in *Fgfr4*^{A/A} mice and wild-type *Fgfr4*^{G/G} mice (Extended Data Fig. 1d–f). Under normal conditions, *Fgfr4*^{A/A} mutant and *Fgfr4*^{G/G} wild-type mice appeared healthy, with no obvious phenotypic differences. However, at the molecular level, primary mouse embryonic fibroblasts (MEFs) derived from embryonic day (E)13.5 embryos of *Fgfr4*^{A/A} and *Fgfr4*^{G/G} mice contained significantly elevated concentrations of key second messenger

signalling molecules, such as ATP and calcium ions (Extended Data Fig. 1g, h). Moreover, quantitative mass spectrometry analyses of four biological replicates identified upregulation of pro-mitotic molecules, including cell cycle and DNA metabolism proteins, in *Fgfr4*^{A/A} MEFs compared with *Fgfr4*^{G/G} MEFs (Extended Data Figs 1i and 2a, b and Supplementary Table 1). Consistent with these results, *Fgfr4*^{A/A} cells displayed an elevated proliferation rate (approximately twofold at 16 h) compared with *Fgfr4*^{G/G} cells (Extended Data Fig. 2c), which was maintained even upon expression of oncogenic signalling molecules, such as KRAS-G12V, BRAF-V600E, CRAF-BxB, EGFR, EGFR-L858R, EGFR-L858R+T790M and EGFR-DEL1 (Extended Data Fig. 2d). Thus, cells expressing endogenous levels of the FGFR4 Arg385 risk variant possess an inherent ability to undergo enhanced proliferation. This is concordant with a cell-intrinsic pro-mitotic role for the *FGFR4* p.G388R germline mutation in multiple cancer types. Intriguingly, although *Fgfr4*^{A/A} cells displayed higher mitogenic properties than *Fgfr4*^{G/G} cells, MAPK/ERK signalling was unchanged (Extended Data Fig. 1e), suggesting that alternative signalling mechanisms are responsible for the increased proliferation.

Interestingly, compared with wild-type cells, surface expression but not total expression of FGFR4 was moderately downregulated both in mouse and in human cells homozygous for the FGFR4 Arg385 and FGFR4 Arg388 alleles, respectively. A surface biotinylation assay pulled down relatively lower amounts of FGFR4 protein from *Fgfr4*^{A/A} MEFs, but it co-precipitated with increased BiP/GRP78 (endoplasmic reticulum chaperone) and tyrosine-sulfated proteins (Golgi specific marker) (Fig. 1a), suggesting folding stress during transmembrane embedding of FGFR4 Arg385 protein variant. BiP/GRP78 is normally upregulated during improper folding of membrane proteins. Consistent with this, flow cytometry staining also detected relatively lower amounts of surface FGFR4 in serum-starved human lung-cancer cell lines (Extended Data Fig. 3a, b) that were homozygous for the rs351855-A allele (Supplementary Tables 2 and 3), regardless of variations in the messenger RNA (mRNA) and total protein expression (Extended Data Fig. 3c, d). Furthermore, double immunofluorescence detection of FGFR4 revealed a significant increase in co-localization of BiP/GRP78 and the FGFR4 Arg385 variant protein in a large number of *Fgfr4*^{A/A} MEFs compared with *Fgfr4*^{G/G} MEFs (Extended Fig. 3e, f). Together, these results suggest that a single amino-acid change from glycine to charged arginine in the transmembrane segment leads to alterations in the kinetics of the FGFR4 receptor maturation process from the endoplasmic reticulum to the cell surface.

Positional analysis of the amino-acid composition of the transmembrane domain (TMD) from different organelles in plants, fungi and vertebrates showed almost no occurrence of arginine residues in transmembrane segments proximal to the cytoplasmic edge^{12,13} (Extended Data Fig. 3g). Intrigued by these reports, we performed bioinformatics analyses encompassing multiple sequence alignments and transmembrane helix predictions on the FGFR4 Arg388 variant protein sequence. The results predicted an alteration in the topology of the

¹Max Planck Institute for Biochemistry, Department of Molecular Biology, Am Klopferspitz 18, 82152, Martinsried, Germany. ²Max Planck Institute for Heart and Lung Research, Molecular Mechanisms of Lung Cancer, Parkstrasse 1, 61231 Bad Nauheim, Germany.

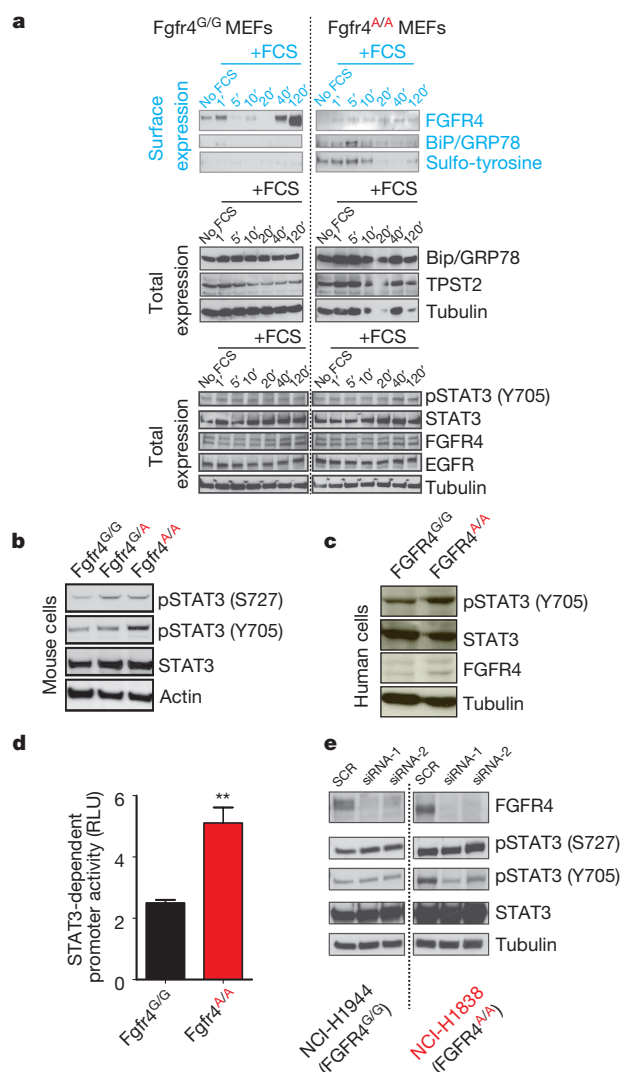


Figure 1 | FGFR4 p.G388R enhances STAT3 activation in mouse and human cells. **a**, Analyses of cell-surface expression of FGFR4, endoplasmic reticulum and Golgi-specific proteins by surface-biotinylation assay in serum-starved MEFs (see Methods). Protein expression on the cell membranes of *Fgfr4*^{G/G} (left) and *Fgfr4*^{A/A} (right) MEFs is displayed in blue and total cell lysate in black. **b**, **c**, Immunoblot analyses of pSTAT3 (Y705) in *Fgfr4*^{G/G}, *Fgfr4*^{G/A} and *Fgfr4*^{A/A} MEFs (**b**) and in human primary breast epithelial cells homozygous for *FGFR4* rs351855-G and *FGFR4* rs351855-A alleles (**c**). **d**, Quantification of STAT3-dependent promoter activity by dual luciferase assay in *Fgfr4*^{G/G} and *Fgfr4*^{A/A} MEFs (means \pm s.e.m., $n=6$, ** $P < 0.01$, unpaired t -test with Welch's correction); RLU, relative light units. **e**, Immunoblot analyses of pSTAT3 (Y705) after knockdown of FGFR4 in H1944 (rs351855-G/G) and H1838 (rs351855-A/A) human lung cancer cells.

transmembrane segment in the FGFR4 Arg388 variant (Extended Data Fig. 3h). We noticed a STAT3 binding motif, YXXQ¹⁴, which was partly hidden in the membrane segment of FGFR4 in all mammalian species (Extended Data Fig. 1a). Furthermore, alignments of the transmembrane segments of all type I membrane proteins in humans suggested that such motifs (YXXQ/C) occurring proximal to inner membrane are fairly common (Extended Data Fig. 4a). Remarkably, such motifs are particularly enriched in the human cluster of differentiation (CD) molecules that are generally considered surface markers for immune cells (Extended Data Fig. 4b, c). An extensive combined analysis of coding region variants using the publicly available SNP data sets obtained from the National Heart, Lung, and Blood Institute (NHLBI) GO ESP Exome Variant Server (<http://evs.gs.washington.edu/EVS/>), Ensembl variation database release 79 (ref. 15) and COSMIC database led to the identification of many similar germline variants that introduce such

cryptic STAT3 binding sites proximal to the membrane. Intriguingly, such germline mutations were found to be co-localized with somatic mutations in the COSMIC cancer data set, calling into question the definitive evidence that they are somatically acquired (Supplementary Table 4). These results, therefore, indicate an as yet undescribed functional relevance of the membrane-proximal STAT3 binding site.

Membrane proteins with short TMDs (17 and 18 amino acids) are largely localized to the endoplasmic reticulum, whereas proteins with long TMDs (21 amino acids) are mainly present at the cell surface¹⁶. Therefore, in light of our results, we hypothesized that if arginine 388 in the FGFR4 Arg388 variant shortened the TMD, the tyrosine in the Y³⁹⁰RGQ³⁹³ motif would be exposed for the binding of cytoplasmic STAT3 proteins. This should sequester STAT3 close to the inner cell membrane, where the major phosphate transfer reactions generally occur, possibly resulting in its activation. Supporting this hypothesis, we observed that STAT3 activation was enhanced in *Fgfr4*^{A/A} MEFs compared with heterozygous or wild-type cells as well as in human *FGFR4*^{A/A} primary breast epithelial cells and cancer cell lines (Fig. 1b, c and Extended Data Fig. 3d). Consequently, this led to at least a twofold increase in STAT3-dependent promoter activity in *Fgfr4*^{A/A} cells, as measured by a dual luciferase STAT3 reporter assay, thereby confirming enhanced STAT3 signalling (Fig. 1d). Furthermore, enhanced STAT3 activation was preserved in immortalized *Fgfr4*^{A/A} MEFs, even upon expression of oncogenic proteins, such as KRAS-G12V, BRAF-V600E, CRAF-BxB, EGFR-WT, EGFR-L858R, EGFR-L858R+T790M and EGFR-DEL1 (Extended Data Fig. 4d, e). Moreover, upon knockdown of STAT3, the enhanced proliferative capacity of *Fgfr4*^{A/A} MEFs was lost. They were indeed more responsive to growth suppression by STAT3 knockdown compared with *Fgfr4*^{G/G} MEFs (Extended Data Fig. 4f, g). Notably, the feedback upregulation of FGFR4 protein levels after STAT3 knockdown was observed only in *Fgfr4*^{A/A} MEFs, thus establishing a strong signalling link between the two (Extended Data Fig. 4f). Similarly, in human cells the FGFR4 Arg388 allele-dependent upregulation of pSTAT3 (Y705) was nullified by knockdown of endogenous *FGFR4* in rs351855-A/A lung cancer cells but not in rs351855-G/G cells (Fig. 1e). Our observation of increased sensitivity of *Fgfr4*^{A/A} MEFs towards inhibition of STAT3 signalling was further corroborated by results from co-cultivation experiments with differentially labelled *Fgfr4*^{A/A} with red fluorescent protein (RFP)- and *Fgfr4*^{G/G} with green fluorescent protein (GFP)-positive MEF cells. Under identical growth conditions, the proliferation rate of *Fgfr4*^{A/A} MEFs was more strongly inhibited than *Fgfr4*^{G/G} MEFs by treatment with inhibitors, such as erlotinib, TGI-101348 and Stattic, but not with MEK inhibitor PD184352 (Extended Data Fig. 5a–c). Similar results were obtained with STAT3 knockdown experiments in co-cultivated MEFs (Extended Data Fig. 5d–g), suggesting that presence of the rs351855-A allele in the germline genome confers increased sensitivity towards inhibition of STAT3 signalling. This indicates dependence on an enhanced basal STAT3 signalling pathway by *Fgfr4*^{A/A} cells. These results present an opportunity to develop personalized therapy for patients with the rs351855 SNP in their germline genome for treatment with STAT3 inhibitors.

To determine whether FGFR4 Arg388 variant-dependent enhanced STAT3 activation is due to a direct interaction with the Y³⁹⁰RGQ³⁹³ motif and not increased tyrosine kinase activity, we devised experiments with truncated FGFR4 constructs (replacing the cytoplasmic segment from the 398th to 802nd amino acids with yellow fluorescent protein (YFP)) and amino (N)-terminally biotin-conjugated transmembrane peptides containing only the eight-amino-acid cytosolic portion. In the transfectants, roughly 50% of FGFR4 Arg388ΔYFP proteins co-localized with STAT3–turquoise (cyan fluorescent protein, CFP) proteins compared with ~10% observed with FGFR4 Gly388 ΔYFP (Extended Data Fig. 6a), and the full-length FGFR4 Arg388 variant co-precipitated with phosphorylated pSTAT3 (Y705) (Fig. 2a). Phosphorylated tyrosine residues serve as docking sites for Src homology (SH2) domains¹⁷ and are crucial for STAT3 binding¹⁴.

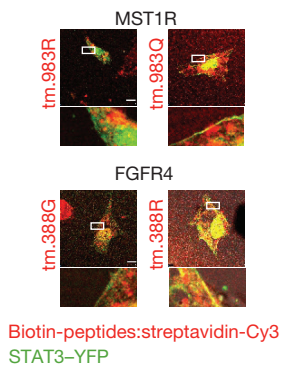


Figure 2 | FGFR4 p.G388R exposes membrane-proximal STAT3 binding site. **a**, Immunoprecipitation of phosphorylated pSTAT3 (Y705) from HEK293T cells transfectants. **b**, Mass spectrometry identification of phosphorylated tyrosine (Y-390) in human FGFR4 Arg388 protein variant (see Methods). Selected peptide-spectrum matches are shown. For ion match table and error map, see Extended Data Fig. 6. N-terminal ions: blue, C-terminal ions: red. **c**, Biotinylated transmembrane peptide sequences representative of membrane-proximal germline variants of MST1R and FGFR4. Transmembrane segment: yellow, SNPs: red. MAF, minor allele frequency. **d**, Co-localization of STAT3 to cell membrane and nucleus induced by transfection of transmembrane peptides with membrane-proximal STAT3 binding sites. Scale bars, 10 μ m. **e**, Immunoblotting for STAT3 in the biotinylated peptide pull-downs from HEK293T cell membrane extracts transfected with indicated peptide sequence variants.

We therefore asked if the exposed tyrosine 390 (Y³⁹⁰) residue in the FGFR4 Arg388 variant is phosphorylated. Indeed, phosphorylation was detectable in the FGFR4 Arg388ΔYFP–His recombinant protein but not the FGFR4 Gly388ΔYFP–His protein, both of which lack the cytoplasmic segment (398–802) (Extended Data Fig. 6b). The modified Y³⁹⁰ was confirmed by mass spectrometry analyses of the purified full-length and truncated FGFR4 variants. The matching phosphorylated Y390-containing peptide identified both in FGFR4 Arg388 and in FGFR4 Arg388Δ is shown (Fig. 2b and Extended Data Fig. 6c).

To further corroborate these findings, we assessed the direct interaction of the FGFR4 Arg388 variant with STAT3 in live cells. We observed a significant increase in the fluorescence resonance energy transfer (FRET) efficiency ratio in cell membranes co-expressing STAT3–turquoise (CFP) and FGFR4 Arg388ΔVenus (YFP) (see Methods) compared with cells co-expressing STAT3–CFP and FGFR4 Gly388ΔYFP ($P = 0.0009$, unpaired t -test). As a reference control, membrane-targeted STAT3 (STAT3 fused with myristoylation signal motif) was co-transfected with FGFR4 Arg388ΔYFP, which exhibited a FRET efficiency ratio of 1 (Extended Data Fig. 7a, b). Importantly,

the FGFR4 Arg388 variant lacking the tyrosine kinase domain was still able to enhance endogenous STAT3 activation (Extended Data Fig. 7c). Next, we gathered additional evidence for a STAT3 binding site proximal to the transmembrane domain using receptor transmembrane segment peptides with and without the YXXQ motif. We synthesized biotin-conjugated transmembrane peptide sequences corresponding to the FGFR4 p.G388R (rs351855) and MST1R p.R983Q (rs375697146) germline variants (Fig. 2c). Upon transfection, a fraction of them reached the cell surface, with the N terminus facing extracellularly, as assessed by flow cytometry analysis of live cell streptavidin-allophycocyanin (APC) binding (Extended Data Fig. 7d); this induced the localization of STAT3–YFP to both the inner cell membrane and nucleus (Fig. 2d). Pull-down of the biotin-conjugated peptides from cell membrane extracts exhibited increased binding of the YXXQ motif-containing peptides to endogenous STAT3 compared with peptides lacking the motif, although the amounts of EGFR associated with membrane extracts containing the wild-type and mutant peptides were similar (Fig. 2e). Finally, both the MST1R-tm.983Q and FGFR4-tm.388R peptides, but not the MST1R-tm.983R and FGFR4-tm.388G versions of the peptides, upregulated STAT3-dependent promoter activity in transfected cells (Extended Data Fig. 7e, f).

Therefore, taken together, these results report a STAT3 signal-regulating function by the germline variants of type I receptor proteins via the membrane-proximal STAT3 binding site, independent of its extracellular or intracellular kinase domains. As a corollary to these findings, an important question that arises is whether mere recruitment of STAT3 to the inner membrane is sufficient to mediate STAT3 phosphorylation. Therefore, to constitutively target STAT3 to the cell membrane, we generated STAT3–turquoise (CFP)-encoding constructs that fused either the myristoylation signal of v-Src¹³ to its N terminus (N1) or the farnesylation and palmitoylation sequence of H-Ras¹⁶ to its carboxy (C) terminus (C1). As controls, we generated unmodified STAT3–turquoise (S) and modified STAT3–turquoise constructs defective in plasma membrane targeting, with inactivating mutations in the myristoylation (N2), farnesylation (C2) and palmitoylation signals (C3), respectively (Fig. 3a). Upon membrane targeting of STAT3 to the plasma membrane, its tyrosine phosphorylation and activation was dramatically increased, as examined by immunoblot analysis of transfected cell lysates (Fig. 3b). The results from dual luciferase STAT3-dependent promoter activity assay performed on *Egfr* wild-type (*Egfr*^{+/+}) and knockout (*Egfr*^{−/−}) MEFs suggested an EGFR-dependent tyrosine phosphorylation of membrane-targeted STAT3 (Fig. 3c). Results from immunoblot analyses of isolated subcellular (membrane, cytoplasm and nucleus) fractions (Fig. 3d) and confocal FRET imaging of the total STAT3 and pSTAT3 (Y705) proteins in transfectants (Extended Data Fig. 8a) all corroborated the membrane localization, increased phosphorylation and nuclear translocation of membrane-targeted STAT3. In the results above, we observed that membrane targeting of STAT3 using the N-terminal fusion was more effective than the C-terminal fusion, which distanced STAT3 from the membrane by a 25 kilodalton CFP. This was interesting, particularly when taking into account the increased affinity of STAT3-β (short isoform lacking C-term 55 amino acids) over STAT3-α towards the FGFR4-tm.388R membrane peptides (Fig. 2e). Further support for this idea was drawn from experiments using plasmids that encoded versions of FGFR4 in which the location of the YXXQ motif in FGFR4 was displaced from the membrane proximity and placed in the juxtamembrane (FGFR4 p.L414Y), tyrosine kinase (FGFR4 p.V550Q) and C-terminal tail (FGFR4 p.L757Q) sequences. Apparently, the pro-mitotic property of FGFR4 was reduced with no increase in pSTAT3 (Y705) levels when the STAT3 binding site was moved away from the membrane proximity, as inferred from immunoblot analyses (Extended Data Fig. 8b) and colony formation assays (Extended Data Fig. 8c). FGFR4 p.Y390A (abolishes membrane-proximal Y³⁹⁰XXQ³⁹³) served as a control. Taken together, these results suggest that the membrane proximity of the STAT3 binding site is crucial for increased tyrosine phosphorylation.

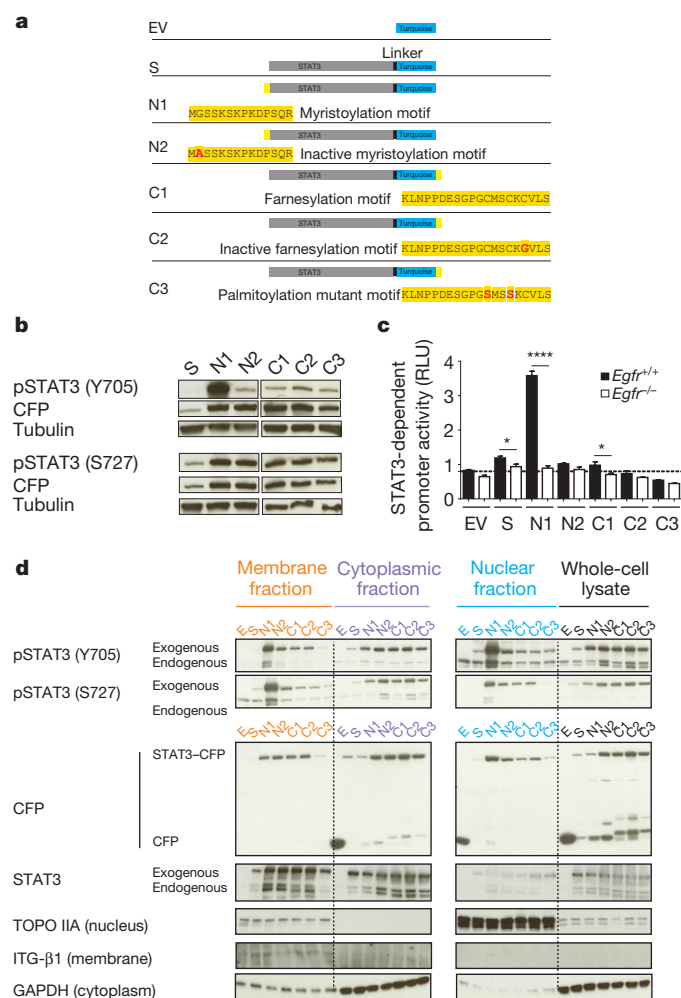


Figure 3 | Targeting STAT3 to cell membrane enhances tyrosine phosphorylation. **a**, Schematic illustration of membrane-targeting STAT3 constructs. Lipid modification signals: yellow box; inactivating mutation: red font. **b**, Immunoblot analyses of pSTAT3 (Y705) in HEK293T transfectants. **c**, STAT3-dependent promoter activity induced by membrane-targeted STAT3 in *Egfr*^{+/+} and *Egfr*^{-/-} MEFs (means \pm s.e.m., $n = 4$; * $P < 0.05$, **** $P < 0.0001$, one-way analysis of variance (ANOVA), Sidak's multiple comparisons test with 95% confidence interval of difference). **d**, Immunoblot analyses in the subcellular fractions (membranes, cytoplasm and nucleus) of HEK293T transfectants. CFP: loading control.

Because similar amounts of EGFR were associated with transmembrane domains (lacking extracellular and intracellular regions) of FGFR4 and MST1R, as inferred from Fig. 2e, we reasoned that EGFR might participate in the tyrosine phosphorylation of STAT3 that was recruited to the inner membrane by the YXXQ motif. While JAK kinase inhibition by ruxolitinib resulted in a complete absence of phosphorylated STAT3 (Y705) expression, similar expression levels of phosphorylated STAT3 (Y705) were observed both in *Fgfr4*^{G/G} and in *Fgfr4*^{A/A} MEFs after 2 h of EGFR inhibition by erlotinib (Extended Data Fig. 9a, b). Thus, specifically blocking the FGFR4 Arg385 variant induced STAT3 phosphorylation. The other tested inhibitors did not restore pSTAT3 (Y705) to normal levels (Extended Data Fig. 9c). Moreover, knockdown of *Egfr* in *Fgfr4* knock-in MEFs resulted in a reduction in pSTAT3 (Y705) levels in *Fgfr4*^{A/A} MEFs (Extended Data Fig. 9d). To further corroborate these results, we transfected *Egfr* knockout and their wild-type counterpart MEFs with FGFR4-truncated constructs and transmembrane peptides. A significant increase both in phosphorylated STAT3 (Y705) expression (Extended Data Fig. 9e) and in STAT3-dependent promoter activity (Extended Data Fig. 9f)

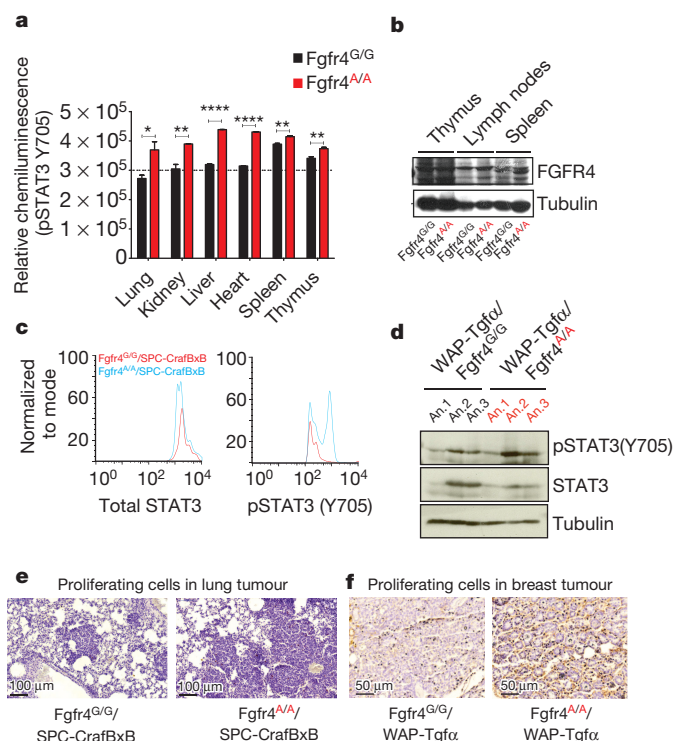


Figure 4 | FGFR4 p.G388R enhances STAT3 phosphorylation in vivo.

a, Quantification of pSTAT3 (Y705) levels in the lung, kidney, liver, heart, spleen, thymus and lymph nodes of *Fgfr4*^{A/A} and *Fgfr4*^{G/G} adult mice (means \pm s.e.m., $n = 5$, * $P < 0.05$, ** $P < 0.01$, **** $P < 0.0001$, two-tailed non-parametric Mann–Whitney rank-sum test). **b**, Immunoblotting for FGFR4 expression in lymphoid organs. **c**, Flow cytometry analyses for pSTAT3 (Y705) expression in lung tumour cells extracted from *Fgfr4*^{G/G};SPC-CrafBxB (red) and *Fgfr4*^{A/A};SPC-CrafBxB (turquoise blue) mouse models for lung cancer. **d**, Immunoblot analyses in breast tumour tissues extracted from *Fgfr4*^{G/G};WAP-Tgfa and *Fgfr4*^{A/A};WAP-Tgfa mouse models for breast cancer. **e**, **f**, Immunohistochemical detection of Ki67⁺ cells, in lung tumours of *Fgfr4*^{A/A};SPC-CrafBxB and *Fgfr4*^{G/G};SPC-CrafBxB mice (5 months old) (**e**) and breast tumours of *Fgfr4*^{G/G};WAP-Tgfa and *Fgfr4*^{A/A};WAP-Tgfa mice (3 months after pregnancy) (**f**). Scale bars, 100 and 50 μ m.

was observed with the FGFR4 Arg388 Δ plasmid and FGFR4-tm.388R peptide in wild-type MEFs but not in *Egfr* knockout cells.

To validate our mechanistic findings *in vivo*, we investigated the basal levels of pSTAT3 (Y705) in whole-organ lysates (Fig. 4a). A significant increase in basal STAT3 activation was apparent in all organs extracted from *Fgfr4*^{A/A} mice. We next analysed lung tumours extracted from *Fgfr4*^{A/A};SPC-CrafBxB and *Fgfr4*^{G/G};SPC-CrafBxB mice (6 months old) and breast tumours from *Fgfr4*^{A/A};WAP-Tgfa and *Fgfr4*^{G/G};WAP-Tgfa mice (3 months after pregnancy). Consistent with results obtained in healthy mice tissues, expression of pSTAT3 (Y705) was increased in the tumour tissues of mice with lung cancer (Fig. 4c) and breast cancer (Fig. 4d). Furthermore, the increase in Ki67⁺ cells in lung tumours of *Fgfr4*^{A/A};SPC-CrafBxB mice (Fig. 4e) and breast tumours of *Fgfr4*^{A/A};WAP-Tgfa mice (Fig. 4f) validated enhanced pro-mitotic signalling within the tumour tissues.

Collectively, our study demonstrates a gain-of-function effect of the cancer-associated rs351855 SNP encoding the FGFR4 Arg388 allele in humans. Given the importance of STAT3 signalling in the immune system and the expression of FGFR4 in lymphoid organs (Fig. 4b), it is plausible that, in addition to promoting growth in malignant cells in a cell-autonomous way, the FGFR4 p.G388R germline mutation may have a cancer-cell extrinsic role in promoting tumour growth *in vivo*. Our continuing investigations in immune cells using knock-in mice suggest a role for the FGFR4 Arg388 variant in suppressing the CD8/T_{reg} lymphocyte ratio (unpublished observations). Our analyses

of publicly available tumour genome data sets revealed additional cell-surface proteins with germline mutations that either create or delete the STAT3 binding site proximal to the membrane (Extended Data Fig. 10 and Supplementary Table 4). However, their association with resistance either to malignancy or to accelerated cancer progression remains to be validated clinically. Our study draws attention to a general need to extend the focus in personalized medicine from heterogeneous tumour-specific somatic mutations to the investigation of patient-specific germline variations.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 August 2014; accepted 14 November 2015.

Published online 16 December 2015.

- Morimoto, Y. *et al.* Single nucleotide polymorphism in fibroblast growth factor receptor 4 at codon 388 is associated with prognosis in high-grade soft tissue sarcoma. *Cancer* **98**, 2245–2250 (2003).
- Fruillanti, E. *et al.* Meta and pooled analyses of FGFR4 Gly388Arg polymorphism as a cancer prognostic factor. *Eur. J. Cancer Prevention* **20**, 340–347 (2011).
- Heinzle, C. *et al.* Differential effects of polymorphic alleles of FGF receptor 4 on colon cancer growth and metastasis. *Cancer Res.* **72**, 5767–5777 (2012).
- Xu, W. *et al.* FGFR4 transmembrane domain polymorphism and cancer risk: a meta-analysis including 8555 subjects. *Eur. J. Cancer* **46**, 3332–3338 (2010).
- Xu, B. *et al.* FGFR4 Gly388Arg polymorphism contributes to prostate cancer development and progression: a meta-analysis of 2618 cases and 2305 controls. *BMC Cancer* **11**, 84 (2011).
- Streit, S., Mestel, D. S., Schmidt, M., Ullrich, A. & Berking, C. FGFR4 Arg388 allele correlates with tumour thickness and FGFR4 protein expression with survival of melanoma patients. *Br. J. Cancer* **94**, 1879–1886 (2006).
- Spinola, M. *et al.* Functional FGFR4 Gly388Arg polymorphism predicts prognosis in lung adenocarcinoma patients. *J. Clin. Oncol.* **23**, 7307–7311 (2005).
- Falvella, F. S. *et al.* FGFR4 Gly388Arg polymorphism may affect the clinical stage of patients with lung cancer by modulating the transcriptional profile of normal lung. *Int. J. Cancer* **124**, 2880–2885, (2009).
- Streit, S. *et al.* Involvement of the FGFR4 Arg388 allele in head and neck squamous cell carcinoma. *Int. J. Cancer* **111**, 213–217 (2004).
- Bange, J. *et al.* Cancer progression and tumor cell motility are associated with the FGFR4 Arg(388) allele. *Cancer Res.* **62**, 840–847 (2002).
- Abecasis, G. R. *et al.* 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
- Sharpe, H. J., Stevens, T. J. & Munro, S. A comprehensive comparison of transmembrane domains reveals organelle-specific properties. *Cell* **142**, 158–169 (2010).
- Buss, J. E., Der, C. J. & Solski, P. A. The six amino-terminal amino acids of p60src are sufficient to cause myristylation of p21v-ras. *Mol. Cell. Biol.* **8**, 3960–3963 (1988).
- Shao, H. *et al.* Structural requirements for signal transducer and activator of transcription 3 binding to phosphotyrosine ligands containing the YXXQ motif. *J. Biol. Chem.* **279**, 18967–18973 (2004).
- Cunningham, F. *et al.* Ensembl 2015. *Nucleic Acids Res.* **43**, D662–D669 (2015).
- Hancock, J. F., Cadwallader, K., Paterson, H. & Marshall, C. J. A CAAX or a CAAL motif and a second signal are sufficient for plasma membrane targeting of ras proteins. *EMBO J.* **10**, 4033–4039 (1991).
- Rotin, D. *et al.* SH2 domains prevent tyrosine dephosphorylation of the EGF receptor: identification of Tyr992 as the high-affinity binding site for SH2 domains of phospholipase C gamma. *EMBO J.* **11**, 559–567 (1992).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the NHLBI GO Exome Sequencing Project and its ongoing studies, which produced and provided exome variant calls for comparison: the Lung GO Sequencing Project (HL-102923), the WHI Sequencing Project (HL-102924), the Broad GO Sequencing Project (HL-102925), the Seattle GO Sequencing Project (HL-102926) and the Heart GO Sequencing Project (HL-103010). We also thank N. Nagaraj for the help with quantitative proteomics and T. W. J. Gadella for providing us pmTurquoise2-N1 and C1 plasmids. We acknowledge T. Mayr for helping us by providing *Fgfr4* knock-in mice for preliminary experiments, and R. Hornberger for technical assistance. We thank M. Sibilia for providing the *Egfr* knockout and wild-type counterpart MEFs. R. Abraham provided FGFR4 extracellular domain-specific anti-human mAb, clone 4FA6#11. We also acknowledge H. Brandstetter, S. Uebel and M. Spitaler for their assistance and services.

Author Contributions A.U. conceptualized and planned the project. A.U. and V.K.U. designed the study. A.U. and V.K.U. coordinated the experimental work and wrote the manuscript with input from co-authors. V.K.U. investigated, analysed and interpreted the results. V.K.U. planned and performed *in vitro* and *in vivo* experiments with assistance from B.S.; normal healthy knock-in mice were generated and analysed by V.K.U.; the knock-in mouse model for lung cancer *Fgfr4^{G/G};SPC-CrafBxB* and *Fgfr4^{A/A};SPC-CrafBxB* was generated and analysed by V.K.U.; U.R.R. provided *SPC-CrafBxB* breeder pairs for this study. The knock-in mouse model for breast cancer *Fgfr4^{G/G};WAP-Tgfα* and *Fgfr4^{A/A};WAP-Tgfα* was generated by B.S. and analysed by V.K.U.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.U. (ulrich@biochem.mpg.de) or V.K.U. (ulaganat@biochem.mpg.de).

METHODS

Plasmids. pCMV-FGFR4 Gly388Δ-Venus, pCMV-FGFR4 Arg388Δ-Venus, pCMV-FGFR4 Gly388Δ-Venus-6×His and pCMV-FGFR4 Arg388Δ-Venus-6×His was constructed by PCR amplification of human FGFR4 (Gly388 variant and Arg388 variant) lacking cytoplasmic domains and inserted into mVenus-N1 between NheI and AgeI sites. The 6×His Tag constructs were generated by PCR-based generation of DNA cassettes 'FGFR4 Gly388Δ-Venus-6×His' and 'FGFR4 Arg388Δ-Venus-6×His' and inserted into mVenus-N1 between NheI and NotI sites. The following primers were used: hFGFR4-DEL-F, 5'-TCTGCTAGCGCCACCATGCGGCTGCTGCTGCCCTGTT-3', hFGFR4-DEL-R, 5'-AGAAGCCGGTCCGCCGTGGAGCGCCTGCCCTC-3'; YFP-HisTag-R, 5'-TCGCGCCGCTTTAATGGTGATGGTGATGATGCTTGTACAGCTCGTCCATGCCGAGA-3'.

Generation of membrane-targeting STAT3 constructs. STAT3-turquoise fusion protein encoding plasmid was constructed by cloning PCR-amplified STAT3 complementary DNA (cDNA) in frame between BglII and SacII sites in pTurquoise2-N1 plasmid. Membrane-targeting STAT3-turquoise fusion constructs were generated by cloning PCR-amplified STAT3 cDNA into pTurquoise2 between BglII and NotI sites. The following primers were used for constructing STAT3 membrane targeting plasmids: N-Memb-STAT3-F, 5'-TCTAGATCTCGCCACC ATGGGCAGCTCCAAATCTAAACCAAGGACCCTTCACAGAGGTCCGG ACTCAGGCTCTAGGCTCAGTGGGAACCAAGCT-3'; N-MyrMut-STAT3-F, 5'-TCTAGATCTCGCCACCAGTGGCACTCCAAATCTAAACCAAGGACC CTTCACAGAGGTCCGGACTCAGGCTCTATGGCTCAGTGGGAACCAAGCT-3'; C-Memb-STAT3-R, 5'-TCTGCGGCGCTCAGGAGAGCACACACTTGCAGC TCATGCAGCCGGGGCCACTCTCATCAGGAGGGTTCAGCTTAGACCTGA GTCCGGACTTGTACAGCTCGTCCATGC-3'; C-PalmMut-STAT3-R, 5'-TCTG CCGCGCTTGTACAGCTCAGCACTTGGAGCTCATGGAGCCGGGGCCAC TCTCATCAGGAGGGTTCAGCTTAGACCTGAGTCCGGACTTGTACAGCT CGTCCATGC-3'; C-FarnMut-STAT3-R, 5'-TCTGCGGCGCTCAGGAGAGCA CACCTTGCAGCTCATGCAGCCGGGGCCACTCTCATCAGGAGGGTTCAG CTTAGACCTGAGTCCGGACTTGTACAGCTCGTCCATGC-3'.

Plasmid construct sequence files. See <http://figshare.com/s/4c05958705c11e5951606ec4b8d1f61>.

T. W. J. Gadella provided pmTurquoise2-C1 and N1.

Transposon-based plasmids, namely ITR-CAG-GFP-ITR and ITR-CAG-DrRed-ITR, were constructed by gateway cloning of GFP and DsRed coding sequence under CAG promoter. Transposase-encoding construct pCMV-SB100 was provided by M.-S. Supprian.

The following plasmids were obtained from Addgene: mVenus-N1, mVenus-C1 and pBabe-puro-KRASV12. The following plasmids were gifted by J. Heuckmann: pBabe-puro-EV, pBabe-puro-EGFR-WT, pBabe-puro-EGFR-L858R, pBabe-puro-EGFR-DEL1. pCdna3-STAT3-YFP was provided by T. Berg. pBabe-puro-CRAF-BxB was provided by U. R. Rapp.

Inhibitors. Small-molecule chemical inhibitors used in this study were as follows: erlotinib (5083S, Cell Signaling), Tarceva (T007500, TRC), ruxolitinib (11609, Cayman Chemical Company), InSolution MEK1/2 Inhibitor III (44968, Calbiochem), InSolution JAK Inhibitor I (420097, Calbiochem), TG101348 (S2736, Selleckchem), Stattic (S7947, Sigma-Aldrich), wortmanin (9951, Cell Signaling) and LY 294002 (1130, Tocris).

Cell lines and medium. Human primary breast epithelial cells and their corresponding culture media were purchased from Zen-Bio and genotyped using primers rs351855-Forw and rs351855-Rev as described in the DNA sequencing section. Lung-cancer cell lines used in this study were as follows: NCI-HCC15, NCI-H520, NCI-H23, NCI-H1944, NCI-H1299, HCC1833, NCI-H1568, NCI-H2126, NCI-H2882, NCI-H1792 and NCI-H358. All cell lines used were obtained from American Type Culture Collection (ATCC) except HCC1833, which was obtained from Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ) and authenticated in-house using a StemElite ID system (Promega, G9530). None of the cell lines used in this study were in the International Cell Line Authentication Committee list of currently known cross-contaminated or misidentified cell lines. Cell lines maintained by our cell bank staff are routinely controlled for mycoplasma contamination. The cell lines used in this study were confirmed to be free of any mycoplasma contamination.

Animal models. All of the experimental protocols were performed as per the Institutional Animal Care and Use Committee at the Max Planck Institute of Biochemistry (www.biochem.mpg.de/en/facilities/animal) and all animal experiments were approved by the Institutional Review Board. In this paper, animal experiments involving breeding and killing for organ extraction came under 'regular animal use' as per the guidelines of Animal Protection Law 2013 (Upper Bavarian Government). All mice used for this study were raised in C57BL/6 background. Normal, healthy *Fgfr4* knock-in mice used were aged 10 weeks. A knock-in

mouse model for non-small-cell lung cancer was generated by breeding *Fgfr4^{ΔA/A}* and *Fgfr4^{G/G}* mice with *SPC-CRAF-BxB* mice. *SPC-CRAF-BxB* induce lung tumours in alveolar type II cells of the lung that can be analysed from 4 months onwards. Six-month-old mice were genotyped and killed for analyses. The knock-in mouse model for breast cancer was generated by breeding *Fgfr4^{ΔA/A}* and *Fgfr4^{G/G}* mice with *WAP-Tgfa* mice as previously described¹⁸. Only female mice 3 months after pregnancy were analysed. Flow cytometry analyses involving cohorts of wild-type and risk-variant groups of mice were done by killing all the mice on the same day. The differences between the means of the wild-type and mutant groups were determined using an unpaired *t*-test. For pSTAT3 (Y705) expression analyses of tumour cells comparing wild-type and risk variant knock-in mice groups, each group consisted of at least five mice, matched for age and gender. All lung cancer and breast cancer mouse models used for tumour analyses were male and females respectively. Tumour-bearing mice were regularly monitored and killed before tumour burden affected their well-being. In the *WAP-Tgfa* transgenic mouse models for breast cancer, spontaneous tumours that arise in mammary pads are visible and measurable. As per our legal institute permit, the maximum tumour volume permitted in *WAP-Tgfa* mouse models of breast cancers was 1,500 mm³ (single tumours); in none of our experiments were these limits exceeded. In the *SPC-CRAF-BxB* transgenic mouse models for non-small-cell lung cancer, the spontaneous tumour *in vivo* does not permit measurement in live animals. However, loss of body weight is proportional to tumour burden. The maximum weight loss permitted as per our animal permit was 10% of the body weight. In none of our experiments with mouse models for lung cancer were these limits exceeded.

Human population genotypes. Allele frequencies for rs2456173, rs1966265, rs376618, rs351855 and rs61737768 were compiled from the 1000 Genomes Project data set release 14 October 2013.

Generation of MEFs. MEFs were generated from E13.5 post-coitum embryos as previously described. In short, embryos from littermates of homozygous and heterozygous genotypes derived from heterozygote parents were separated from placenta and embryo sac. Head and red organs were dissected out and the remaining tissue was finely minced and passed through a cell strainer before washing them in PBS. Two weeks after cultivation, cells were immortalized using equal amounts of transposon-based SV40-T antigen-encoding plasmids. Cells were cultivated for at least a month before preparing frozen stocks and using for downstream experiments.

Surface biotinylation assay. Biotinylation of cell-surface proteins (biotinylation of extracellular exposed domains) in serum-starved knock-in MEFs followed by precipitation of avidin-bound proteins was achieved using cell-impermeable EZ-Link Sulfo-NHS-LC-Biotin (Pierce/Thermo Scientific, 21335) by following the manufacturer's protocol. Briefly, at indicated time points after 10% FCS addition, MEFs were washed with ice-cold PBS (pH 8.0) to remove amine-containing media and proteins from cells. Sulfo-NHS-LC-Biotin reagent (2 mM) was added and incubated for varying time points on ice. Labelled cells were washed three times in PBS containing 100 mM glycine and pellets were lysed in lysis buffer and divided into two parts. One part was used for precipitation of biotinylated proteins and the other for probing the total cellular amounts of proteins. Biotinylated proteins were pulled down using Avidin beads (Pierce, 20219) and probed by immunoblot experiments.

Retroviral transduction. To stably express retroviral plasmids in MEFs, Phoenix-Ecotropic retroviral packaging cell lines were used. Two days after transfection using Lipofectamine 2000 (Life Science Technologies, 11668027), cell culture supernatants were centrifuged at 1,200 r.p.m (130g) for 3 min and filtered using a 0.45 μm filter. MEFs were transduced with retroviral particles containing supernatants using ViraDuctin Retrovirus Transduction Reagent (Cell Biolabs, RV200) as per the manufacturer's protocol.

Phospho-STAT3 (Y705) sandwich enzyme-linked immunosorbent assay. Organs were extracted from indicated mice and lysed in 1× cell lysis buffer (Cell Signaling, 9803) using a tissue homogenizer (IKA T-18, Ultra Turrax). Lysates quantified by bichinchoninic acid (BCA) reagent (Pierce, 23225) for equal protein amounts were used the assay. Levels of phosphorylated STAT3 (Y705) in organ lysate was measuring using a PathScan Phospho-STAT3 (Y705) Sandwich ELISA kit (Cell Signaling, 7149C, 7149) as per the manufacturer's instructions. The data in Fig. 4a represent the relative chemiluminescence light units. The measurements were repeated three times with the same animal lysates using a chemiluminescence kit (Cell Signaling Technology, 7149) and twice by measuring the absorbance of organ lysate preparations from additional mice (three mice per group) using a spectrophotometric kit (Cell Signaling Technology, 7300). The figure represents the data from one chemiluminescence-based measurement assay.

Purification of recombinant proteins. Truncated human FGFR4 (1-397) fused in-frame to YFP-6×His was inserted into pEYFP (Clontech) vector and expressed in HEK293E-EBNA1 (MPI core facility) cell lines. Seventy-two hours after transfection, cells were lysed in binding buffer (20 mM sodium phosphate,

500 mM sodium chloride, 40 mM imidazole, 6 M urea, pH 7.4) containing PhosSTOP tablets (Roche, 04906837001). The lysate was treated with DNase (Thermo Scientific, EN0521) 20 µg ml⁻¹ and Benzonase followed by sonication. The cell lysate was loaded on to HisTrap FF crude 5 ml column in binding buffer. Elution was done in a one-step procedure using elution buffer (20 mM sodium phosphate, 500 mM sodium chloride, 500 mM imidazole, urea, 6 M pH 7.4). Full-length C-term GFP-tagged human FGFR4 variants were purified using a Miltenyi Epitope-tagged protein isolation kit and adapted large-scale samples with a Midi-MACS column.

Immunofluorescence microscopy. Cells were grown overnight in ultraviolet-sterilized glass slides in culture dishes. After plating, cells were washed in PBS and fixed in 4% paraformaldehyde-PBS at room temperature (22–24°C) for 15 min. Permeabilization was done in 0.2% Triton X-100 in PBS for 20 min. Primary antibodies were incubated overnight at 4°C and secondary antibodies for 45 min at room temperature. Co-localization rate was calculated as the ratio of area of the co-localized fluorescence signals to an area of the image foreground.

FRET analysis. HEK293T cells were co-transfected with either pCMV-hFGFR4 Gly388-CFP and pCMV-STAT3-YFP or pCMV-hFGFR4 Arg388-YFP and pCMV-STAT3-CFP DNA in equal amounts by the method of reverse transfection in glass-bottomed culture dishes (35 mm high) (Ibidi, 81156). For FRET localization of STAT3 phosphorylation, pCMV-STAT3-turquoise, pCMV-N1-STAT3-turquoise, pCMV-N2-STAT3-turquoise, pCMV-STAT3-turquoise-C1, pCMV-STAT3-turquoise-C2 and pCMV-STAT3-turquoise-C3 constructs were used. Two days after transfection, imaging was performed either in a spinning disc confocal microscope (PerkinElmer UltraVIEW vox) or in an sp8 Leica confocal microscope at a controlled temperature of 37°C and 5% CO₂ conditions. For CFP images, the light path was set to excite the sample at 2% power from a 405 nm laser and emission was collected from 454 to 568 nm. For YFP images, the light path was set to excite the sample at 2% power from 516 to 621 nm. For positive control, CFP fused to YFP was cloned. After correcting for emission cross-talk and background intensity, the FRET/CFP ratio was calculated. Analyses were performed using an sp8 Leica TCS FRET sensitized emission application. FRET efficiency was calculated using the following method:

$$E_A(i) = \frac{B - A \times \beta - C \times (\gamma - \alpha \times \beta)}{C \times (1 - \beta \times \delta)}$$

where *A*, *B*, *C* correspond to the intensities of the three signals (donor, FRET, acceptor) and α , β , γ and δ are the calibration factors generated by acceptor only and donor only references. The ratiometric calculation $E = B/A$ is used in samples with a fixed stoichiometry (1:1) of donor and acceptor.

Confocal image files source data. See <http://figshare.com/s/66530d9e705c11e5a00c06ec4b8d1f61>.

Calcium assay. Intracellular calcium levels were measured using Fluo-4 Direct calcium assay reagent (Invitrogen, F10471). MEFs were detached using 1 mM EDTA and washed in PBS. Cells were incubated in serum-free medium containing Fluo-4 direct reagent 1 × for 20 min and washed afterwards in PBS. The fluorescence signal was measured in a flow cytometer (FACS Calibur) using an argon laser at FL1 green channel.

ATP assay. The relative levels of ADP and ATP were measured using an ApoSENSOR ADP/ATP Ratio Assay kit (BioVision, Cat K255-200) in the risk variant knock-in *Fgfr4^{A/A}* MEFs, *Fgfr4^{G/G}* heterozygous MEFs and wild-type counterpart *Fgfr4^{G/G}* MEFs. MEFs were lysed in 1 × cell lysis buffer (Cell Signaling, 9803) and total protein was estimated using BCA reagent (Pierce, 23225). Lysates quantified for same protein amount by BCA assay were mixed with equal amounts of nucleotide releasing buffer and incubated at room temperature for 5 min. To the prepared sample lysate, ATP monitoring enzyme was added to a volume of 10% of total volume. Luminescence indicating ATP level was measured after about 2 min in a luminometer (EG&G Berthold Technologies, LB96v). To measure the ADP level, ADP converting enzyme was added to 1% of total volume and the luminescence measured after about 2 min.

Western blot analysis. Whole-cell lysates were prepared using 1 × cell lysis buffer (Cell Signaling, 9803) containing cComplete, mini, EDTA-free tablets (Roche 11836170001) and PhosSTOP tablets (Roche 04906837001). Equal concentrations (20–30 µg) were loaded after a (BCA) assay, were run out on 4–15% Mini-PROTEAN TGX Gels (Bio-Rad 456-1083) and subsequently transferred onto a nitrocellulose membrane. The blots were blocked in 1 × NET-Gelatin buffer (1.5 M NaCl, 0.05 M EDTA, 0.5 M Tris pH 7.5, 0.5% Triton X-100 and 0.25 g ml⁻¹ gelatin) and incubated with primary antibodies overnight at 4°C. Fractions of cell membranes were prepared using a FOCUS Membrane Protein Kit (G Biosciences, 786-249); cytoplasm and nucleus were prepared using Nucleus (Novagen, 71183-3).

Antibodies for western blotting. The following antibodies were used for western blotting. Rabbit anti-FGFR4 (Cell Signaling, 8562), Rabbit anti-ERK1/2

(Cell Signaling, 4695S), Rabbit anti-pERK1/2 (Cell Signaling, 4376S), Rabbit anti-FGFR4 (Santa Cruz, H121, sc9006), Mouse anti-BrdU (Cell Signaling, 5292), Rabbit anti-pSTAT3 (Y705) (Cell Signaling, 9145), Rabbit anti-pSTAT3 (S727) (Cell Signaling, 9134), Rabbit anti-STAT3 (Cell Signaling, 4904), Mouse anti-STAT3 (Cell Signaling, 9139), Mouse anti-BiP/GRP78 (BD Transduction Labs, G73320-050), Rabbit anti-BiP/GRP78 (Abcam, ab21685), Rabbit anti-ITGβ1 (Cell Signaling, 4706), anti-sulfonyltyrosine (Millipore, 05-1100), Rabbit anti-TPST2 (Abcam, ab157191), Mouse anti-EGFR (BD Transduction Labs, E12020), Rabbit anti-pJAK2 (Cell Signaling, 3776), Rabbit anti-JAK2 (Cell Signaling, 3230), Rabbit anti-BRAF (Santa cruz, sc-9002), Rabbit anti-MEK1 (Cell Signaling, 12671), Mouse anti-phospho Tyrosine (Cell Signaling, 9411), Mouse anti-VSV tag (Home Made), Rabbit anti-HIS tag (Cell Signaling, 2365S), Mouse anti-GFP (Home Made), Mouse anti-Tubulin (Sigma, 9026), Rabbit anti-GAPDH-HRP (Cell Signaling, 8884), horseradish-peroxidase-conjugated secondary antibodies and an ECL kit (GE Healthcare/Amersham Pharmacia Biotech, 32106) were used to detect protein signals. Multiple exposures were taken to select images within the dynamic range of the film (GE Healthcare Amersham Hyperfilm ECL, 28906838). Normalization was done using tubulin bands.

Immunoprecipitation. Transfectants were lysed in 1 × cell lysis buffer (Cell Signaling, 9803) containing cComplete, mini, EDTA-free tablets (Roche 11836170001) and PhosSTOP tablets (Roche 04906837001). Lysates were cleared and incubated with primary antibody overnight at 4°C. Dynabeads Protein A (10006d, Life Technologies) (50 µl) was added per sample and incubated with rocking for an additional 4 h. Magnetic-bead-bound proteins were separated using a DynaMag-2 magnet (12321, Life Technologies). After five washes, co-immunoprecipitated proteins were extracted in 3 × Laemli Buffer. For samples from peptide transfectants, a Dynabeads Streptavidin Kit (65801D, Life Technologies) was used.

Flow cytometry. Surface staining for FGFR4 in human cancer cell lines was performed using custom-generated monoclonal antibody raised against extracellular segments of human FGFR4 (U3-Daiichi Sankyo, Clone 4FA6).

Single-cell suspensions of lung tumours were prepared for staining. Erythrolysis was performed by ACK lysis buffer (1.5 M NH₄Cl, 100 mM KHCO₃, 10 mM EDTA-Na₂, pH 7.4). Tumours were first sliced into small pieces and resuspended in 10 ml of digestion cocktail (0.03 g of Liberase Thermolysin Medium (Roche, 05401119001) and 1.3 mg of DNase I (Thermo Scientific, EN0521)) reconstituted in RPMI complete medium. Digestion was performed with gentle agitation at 37°C for 30 min.

Single-cell suspensions were stained with the following antibodies: Rabbit anti-pSTAT3 (Y705) (Cell Signaling, 9145) and Rabbit Isotype Control (Cell Signaling, 3900). Goat anti-Rabbit-APC (Dianova, 111-136-144) was used as secondary detection antibody. Data were analysed using FloJo software version 10.0.7.

Enzyme-linked immunosorbent assay. IL-10 levels in equal volumes of mouse serum samples were quantified using mouse IL-10 ELISA ready-set-go kits (ebioscience, 887104-22) by following the manufacturer's instructions.

Immunohistochemistry and immunofluorescence. Tissues were fixed overnight in 4% paraformaldehyde in PBS (pH 7.4) at 4°C. Fixed tissues were embedded in paraffin and sliced. Sections were prepared for staining first by deparaffinization followed by hydration in the following solutions: three washes of xylene for 5 min each; two washes of 100% ethanol for 10 min each; two washes of 95% ethanol for 10 min each; and two washes in distilled water for 5 min each. Antigen retrieval was obtained by incubation with heated citrate buffer (sodium citrate 10 mM, pH 6) for 15 min. Immunohistochemistry was performed as per the standard procedures. Briefly, after antigen retrieval, sections were incubated 3% hydrogen peroxide for 10 min to quench endogenous peroxidase activity. Non-specific background staining was blocked by incubating in UltraVision Block (Thermo Scientific, TA-060-PBQ) for 5 min at room temperature. Ki67 staining was done by incubating in Rabbit anti-Ki67 mAb (Cell Signaling, 9027) at a dilution of 1:400 overnight at 4°C. Detection was achieved using HRP Polymer (Thermo Scientific, TL-060-PH) followed by incubation with peroxidase-compatible DAB chromogen.

For immunofluorescence, anti-Mouse CD8a-FITC, clone 53-6.7 (eBioscience, 11-0081-82) was used.

Real-time PCR with reverse transcription. Total RNA was isolated using an RNeasy Kit (Qiagen, 74104). RNA was reverse transcribed into cDNA by random hexamer with a First Strand cDNA Synthesis Kit (Thermo Scientific, K1622). A StepOne Plus Real Time PCR System (Applied Biosystem) and Fast SYBR Green Master Mix (Life Science Technologies, 4385612) were used for quantitative RT-PCR. Primers used were as follows: mouse *Fgfr4* (forward: 5'-CAAGTGGTTCGTGCAGAGG-3'; reverse: 5'-CTTCATCACCTCCATCTCGG-3') and *Hprt* (forward: 5'-CTTCCTCCTCAGACCGCTTT-3'; reverse: 5'-TTTCCAAATCCTCGGCATA-3').

Transcript levels in human cell lines were quantified by the method of Taqman real-time PCR using reagents and probes from Integrated DNA Technologies. The primers and probes used were as follows: human *FGFR4* (forward:

5'-TTCTCACAGCTCTCAGGGA-3'; reverse: 5'-CAGGTGAGCAGGACCCCT-3'; probe: 5'-FAM-CAGGCTCGA(ZEN)GGAAGGCAGTTGG3IABkFQ3'; human *HPRT1* (forward: 5'-GTATTCATTATAGTCAAGGGCATATCC-3'; reverse: 5'-AGATGGTCAAGGTCGCAAG-3'; probe: 5'-FAM-TGGTGAAAA(ZEN)GGACCCACGAAGT3IABkFQ-3').

Mass spectrometry. MEFs were cultivated in 15 cm dishes; four biological replicates that were serum-starved overnight were directly lysed on the culture dish using the guanidinium hydrochloride protein digestion method and were subjected to a total proteome analysis by mass spectrometry. The lysates were then sonicated and heated briefly before dilution followed by sequential digestion with LysC and trypsin proteases in solution. After overnight digestion, the peptides were purified with StageTips and measured on a benchtop Orbitrap mass spectrometer as described elsewhere. All raw files were processed using MacQuant software and bioinformatic analysis of quantitative results was performed using the Perseus platform.

For identification of modified tyrosine in human FGFR4 Arg388-GFP, full-length recombinant proteins were purified from HEK293E-EBNA1 cell transfectants by using His-Trap crude chromatography columns in buffers containing phosphatase inhibitors. Two independent eluates were prepared for mass spectrometry analyses. Purified recombinant proteins were resolved in 10% SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and in-Gel digestion procedures were adopted. Sequential digestion was performed with LysC and GluC overnight. Graphical representations of the selected peptide-spectrum matches are shown in Extended Data Figures and source data. The ion table in the bottom panel shows the calculated mass of possible fragment ions. If a fragment ion is found in the spectrum, its mass value is displayed in colour. N-terminal ions are shown in blue and C-terminal ions are shown in red. The 'error map' shows the mass errors of matched fragment ions. The m/z ratio is displayed on the x axis and the error on the y axis in daltons. Each matched fragment ion is represented by a dot. A fragment ion is found if the relative intensity of the matching peak is at least 2%. Samples were analysed using a Q Exactive Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Scientific). Raw files were processed and analysed using the PTM module of PEAKS 7 software (BSI).

Dual luciferase reporter assays. Mouse embryonic fibroblast cells (cultured at 37°C, 7% CO₂ in RPMI containing 4.5 g l⁻¹ glutamine and 10% FBS) were transiently transfected for 48 h with reporter plasmids Cignal STAT3 Reporter (luc) kit (Qiagen, CCS9028L) and pTk-Renilla using Lipofectamine 2000 (Life Science Technologies, 11668027). Similarly, HEK293T cells (cultured at 37°C, 7% CO₂ in RPMI containing 4.5 g l⁻¹ glutamine and 10% FBS) were transiently transfected for 48 h with reporter plasmids Cignal STAT3 Reporter (luc) kit (Qiagen, CCS9028L), pTk-Renilla and plasmids pCMV-hFGFR4-388GlyDel-YFP and pCMV-hFGFR4-388ArgDel-YFP using Lipofectamine 2000 (Life Science Technologies, 11668027). Luciferase was measured with the Dual Glo Luciferase Assay System (Promega, E1910) according to the manufacturer's instructions. Briefly, Dual Glo Luciferase Reagent was added to the cells and, after incubation for 10 min, firefly luciferase activity was measured with a luminometer (EG&G Berthold Technologies, LB96v). Reactions were stopped by treatment for 10 min with Dual-Glo Stop and Glo Reagent and renilla luciferase activity was then measured. For some samples where increased sensitivity was required, a Nano-Glo Dual Luciferase Reporter Assay Prototype Kit (Promega, N1110) was used.

DNA, siRNA and peptide transfection. DNA and peptide transfection was done using Lipofectamine 2000 (Life Science Technologies, 11668027) as per the manufacturer's instructions.

Success of peptide delivery inside the cells was evaluated by flow cytometry assessment of surface and intracellular biotinylated peptides. For surface quantification, PBS-washed cells were fixed in 2% PFA 16 h after peptide transfection. Levels of biotinylated peptide on cell surfaces were then assessed using streptavidin-APC. Untransfected cells were used as negative controls. For intracellular quantification, 4% paraformaldehyde-fixed cells were permeabilized in BD permeabilization buffer and levels were assessed using streptavidin-APC.

RNA transfection used RNAi Max (Life Science Technologies, 13778-150) by following the manufacturer's guidelines. The following siRNAs were purchased from Cell Signaling: control siRNA 6568S, *Fgfr4* siRNA-1 12472S, *Fgfr4* siRNA-2 12669S, *Stat3* siRNA-1 6353, *Stat3* siRNA-2 6353. Mouse-specific

ON-TARGETplus *Egfr* siRNA SMART pool was purchased from Dharmacon (L-MOUSE-XX-00-0529123373).

BrdU incorporation assay. For flow-cytometry-based detection of BrdU incorporation, serum-starved MEFs were incubated in RPMI culture medium containing 1 × BrdU. After indicated time points, cells were detached and fixed in 4% paraformaldehyde in PBS for 30 min at room temperature. Permeabilization was done by incubating with ice-cold 100% methanol for 20 min at 4°C followed by washing in PBS. Mouse anti-BrdU antibody (Cell Signaling, 5292) or mouse isotype control was added and incubated overnight at 4°C. Allophycocyanin-conjugated anti-mouse (Dianova, 115-136-146) was used as secondary antibody and measured in a flow cytometer.

For measuring BrdU incorporation in multiple samples in 96-well plates, a BrdU Cell Proliferation Assay Kit (Cell Signaling, 6813) was used and the manufacturer's instructions followed.

DNA sequencing. Genomic DNA from mouse tissues and human lung-cancer cell lines was isolated using a DNeasy Blood and Tissue Kit (Qiagen, 69506). PCR was done in Q5 High Fidelity 2 × Master Mix (NEB, M0492L) and amplicons obtained using primers rs351855-Forw (5'-CACATATGTTGGGAGCTGGGAG-3') and rs351855-Rev (5'-CTGCAAAGTGGGAGACTTGG-3') were sent for in-house sequencing. DNA samples were sequenced by the microchemistry core-facility using an ABI 3730 sequencer and BI Big Dye 3.1 sequencing chemistry. The following sequencing primers were used to genotype rs351855 (c.1162G>A): hF4-TMs-MvaI (5'-GACCGCAGCAGCGCCCGAGGC-3') and hF4-TMas-MvaI (5'-AGAGGGAAGCGGGAGAGCTTCTG-3'). The sequence was analysed in FinchTV version 1.4.0 (Geospiza) and Seqman Pro (DNASTAR Lasergene 12 Core Suite). Raw sequencing files and contig assembly files are deposited in 'figshare'.

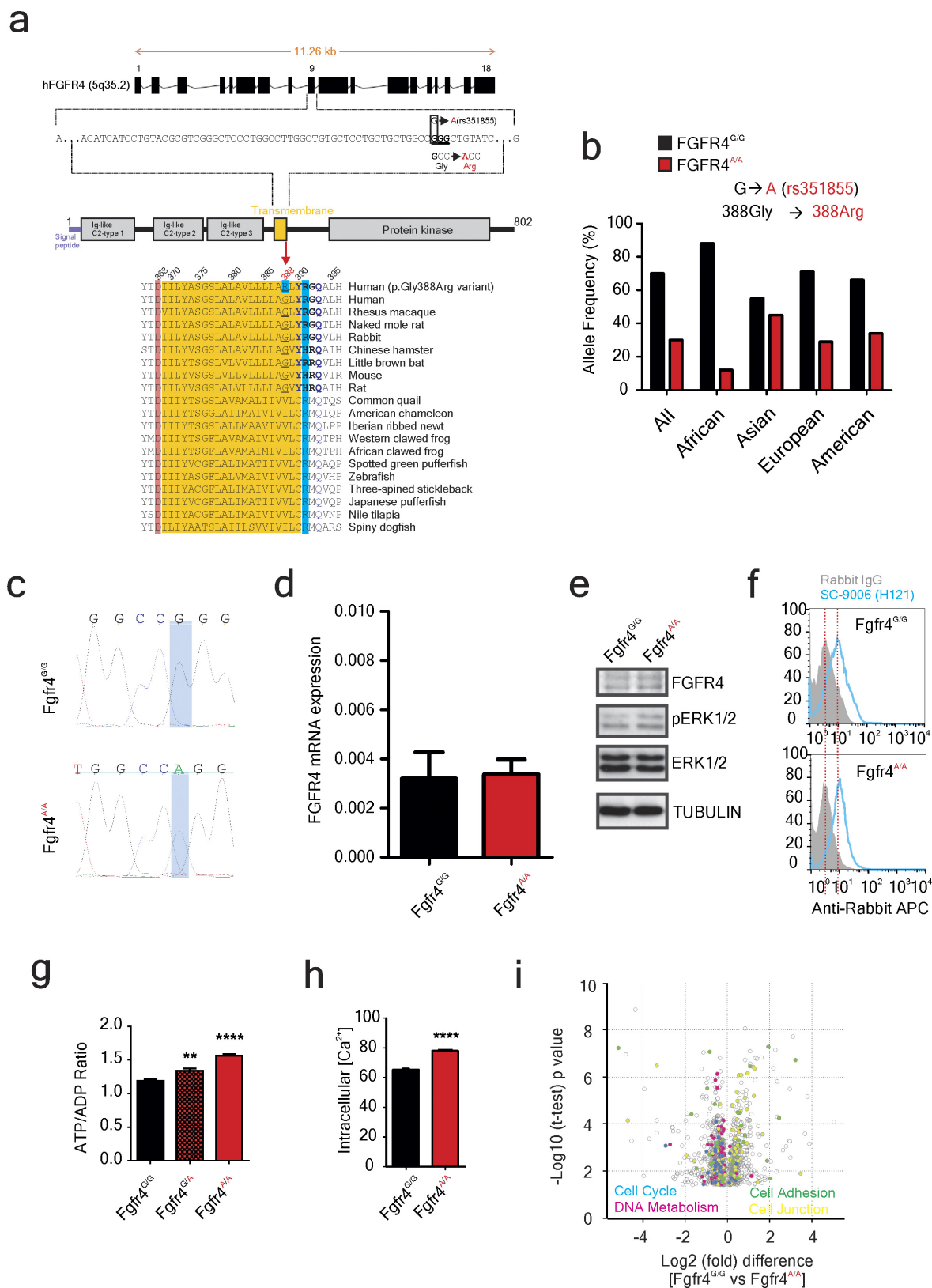
Time-lapse video microscopy. MEFs transfected with pCMV-hFGFR4 Gly388-Venus and pCMV-hFGFR4 Gly388-Venus were imaged using a spinning disc confocal microscope (PerkinElmer UltraVIEW vox) in controlled temperature (37°C) and CO₂ (5%) conditions. A time-lapse of 30 s was fixed and imaged for 1 h. Images were analysed in Volocity software (PerkinElmer). Raw image files including metafiles are deposited in 'figshare'.

Soft-agarose colony formation assay. A colony formation assay was performed using cells transfected with the indicated plasmid grown under plasmid-specific antibiotic selection for 3 weeks in 24-well plates. Spheroid colonies of sizes greater than 80 µm were counted under a ×10 objective.

Statistical analysis. Statistical analyses were performed using Prism software (GraphPad Prism). To detect substantial effects between wild-type and mutant variants of FGFR4, sample sizes were chosen on the basis of standard deviation in the measurements under given experimental conditions. The sample size calculations were determined as per the recommendations of ref. 19. Biological and measurement replicates are indicated in the corresponding figure legends and statistical methods. For immunohistochemical analyses of tumours, a minimum of five mice in a group of age- and gender-matched littermates were used. Animals from each litter were randomly chosen for tumour extraction, and experiments were performed by a co-author unaware of the genotypes. All *in vitro* and immunoblots were performed by a co-author unaware of the treatment or outcome until the end. For statistical analyses, in general for two-group comparisons, we used a Mann-Whitney rank-sum test or unpaired *t*-test with Welch's correction. For multiple group comparisons, two-way ANOVA with Sidak's or Tukey's comparison test was used. All *P* values are two-tailed; the criterion for statistical significance was *P* < 0.05. Values of *P* < 0.05, *P* < 0.001 and *P* < 0.0001 are denoted by *, ** and *** respectively. All data are represented as means either ± s.d. or ± s.e.m.

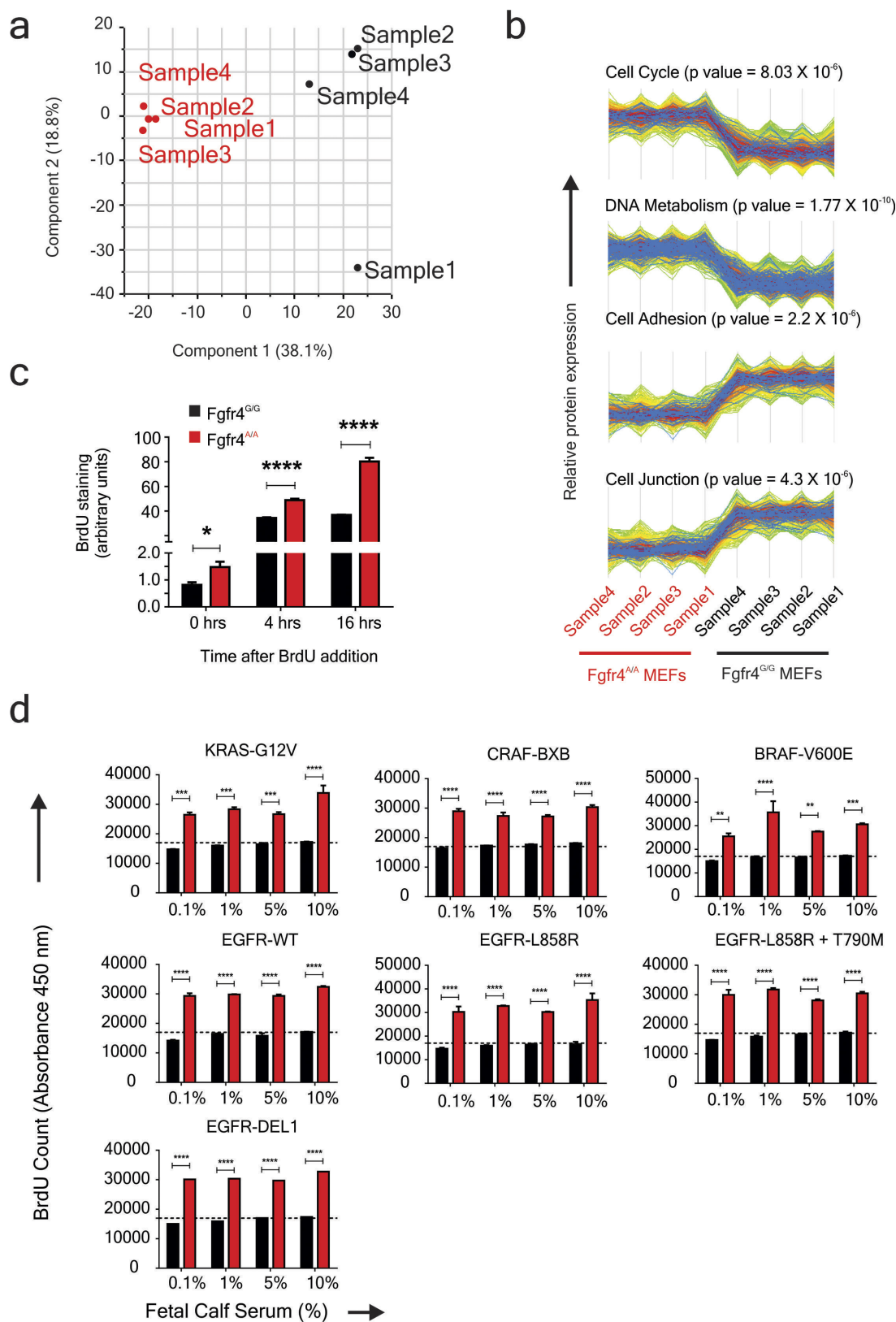
Accession codes. The mass spectrometry proteomics data have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository under data set identifier PXD00313 (ref. 20).

- Sandgren, E. P., *et al.* Inhibition of mammary gland involution is associated with transforming growth factor α but not *c-myc*-induced tumorigenesis in transgenic mice. *Cancer Res.* **55**, 3915–3927 (1995).
- Parker, R. A. & Bermann, N. G. Sample size: more than calculations. *Am. Stat.* **57**, 166–170 (2003).
- Vizcaino, J. A., *et al.* ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nature Biotechnol.* **32**, 223–226 (2014).



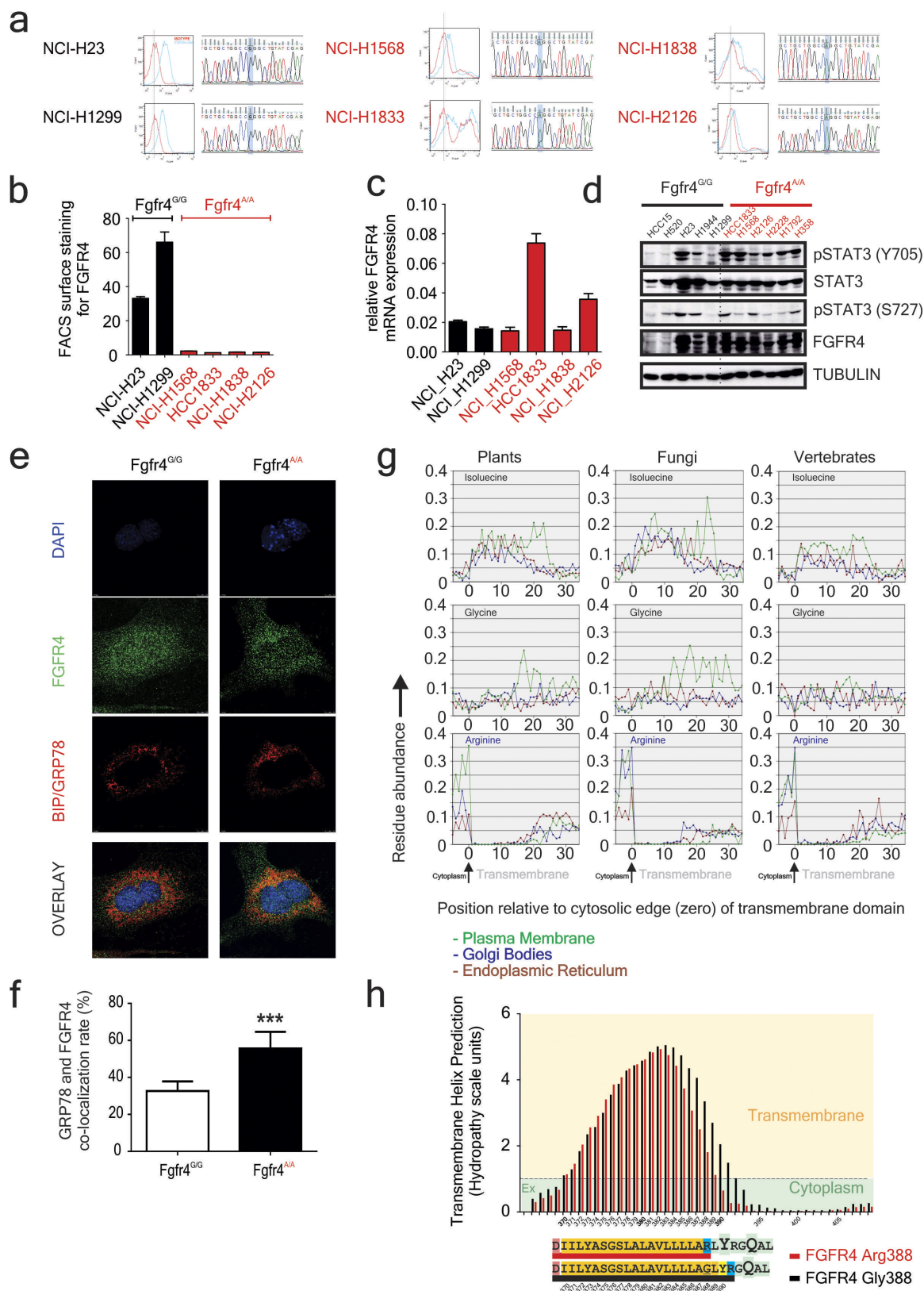
Extended Data Figure 1 | Effect of the rs351855 SNP in the knock-in MEFs. **a**, Human *FGFR4* gene structure depicting the single nucleotide change (rs351855) from guanine to adenine in exon 9 (red arrow). The rs351855-G (ancestral) allele is conserved across mammals, whereas the rs351855-A allele 388 occurs in approximately 30% of humans (according to the 1000 Genomes Project data set). Hydrophobic aliphatic 388Ile/Val evolved to charged 388Arg. **b**, Histogram showing the frequencies of the rs351855-G/A (p.Gly388Arg) allele in humans, including Africans (AFR), Asians (ASN), Europeans (EUR) and Americans (AMR), according to the data obtained from the 1000 Genomes Project (released 14 October 2013).

c–e, SNP sequence (**c**), transcript levels (**d**) and immunoblot analysis (**e**) for FGFR4, pERK1/2, ERK1/2 and tubulin, and, **f**, intracellular flow cytometry staining for FGFR4 expression in *Fgfr4*^{G/G} and *Fgfr4*^{A/A} MEFs. **g**, Relative levels of ADP and ATP in the *Fgfr4*^{A/A}, *Fgfr4*^{G/A} and *Fgfr4*^{G/G} MEFs (means \pm s.d., $n = 10$, ** $P < 0.01$, **** $P < 0.0001$, two-tailed unpaired t -test, 95% confidence level). **h**, Intracellular free calcium levels in *Fgfr4*^{A/A} and *Fgfr4*^{G/G} MEFs (mean \pm s.d., $n = 5$, **** $P < 0.0001$). **i**, Comparison of the total proteome of *Fgfr4*^{G/G} and *Fgfr4*^{A/A} MEFs by quantitative mass spectrometry (see Methods). Cell-cycle proteins, blue; DNA metabolism proteins, pink.



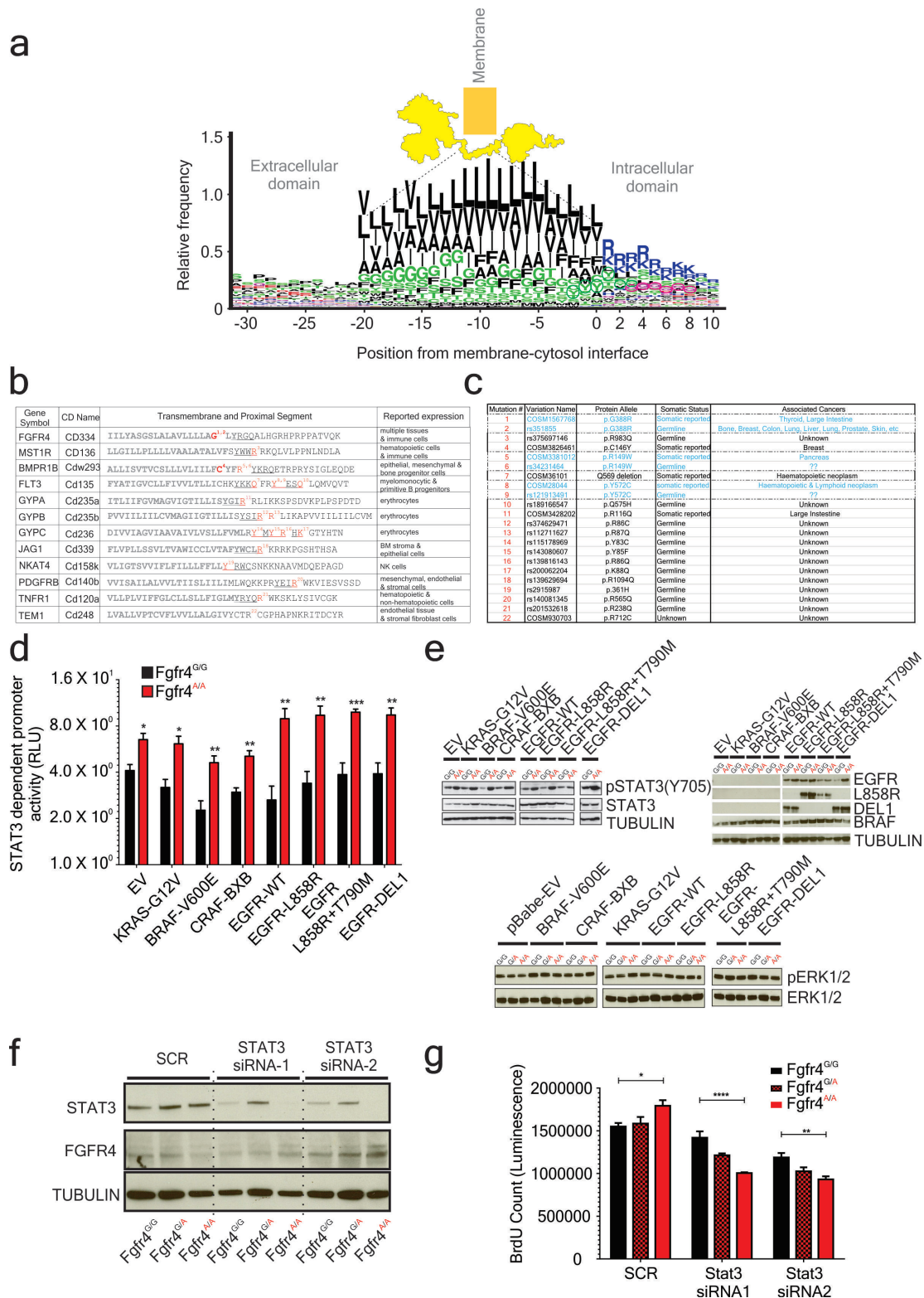
Extended Data Figure 2 | Effect of rs351855 SNP expression on cell proliferation. **a**, Principal component analysis of the total proteome of quadruplicate samples of Fgfr4^{G/G} and Fgfr4^{Δ/Δ} MEFs. **b**, A line plot of the quantitative mass spectrometry data shows the differentially regulated proteins belonging to four gene ontology categories, including cell cycle, DNA metabolism, cell adhesion and cell junction (cut off \log_2 (fold change) ± 0.5 , Benjamini–Hochberg-corrected $P < 0.05$). **c**, Quantification of proliferation in Fgfr4^{Δ/Δ} and Fgfr4^{G/G} MEFs (see Methods) (mean \pm s.d.,

$n = 5$, * $P < 0.05$, **** $P < 0.0001$, two-tailed unpaired t -test with a 95% confidence level). **d**, The proliferation of immortalized MEFs derived from Fgfr4^{G/G} and Fgfr4^{Δ/Δ} mice that were stably transduced with retroviruses encoding human oncogenes, including KRAS-G12V, BRAF-V600E, CRAF-BxB, EGFR-WT, EGFR-L858R, EGFR-L858R+T790M and EGFR-DEL1 (means \pm s.d., $n = 5$, using two-way ANOVA and Sidak's multiple comparisons test).



Extended Data Figure 3 | Effect of the rs351855 SNP on the transmembrane segment of FGFR4. **a**, The surface expression of FGFR4 was detected using a homemade mouse anti-FGFR4 mAb that detects the extracellular domain of FGFR4 by flow cytometry staining. Isotype control: red; FGFR4 staining: turquoise blue. Rs351855 SNP genotyping data. **b**, Quantification of relative surface expression levels of FGFR4. **c**, *FGFR4* mRNA expression. *HPRT1* as internal standard. Black: *FGFR4*^{G/G}; red: *FGFR4*^{A/A} (means \pm s.d., $n = 3$). **d**, Immunoblot analyses for total FGFR4 and pSTAT3 (Y705) expression in human lung cancer cell lines. Black: *FGFR4*^{G/G}; red: *FGFR4*^{A/A}. **e**, Co-localization analyses of FGFR4 and endoplasmic reticulum chaperone BiP/GRP78 proteins. FGFR4: green;

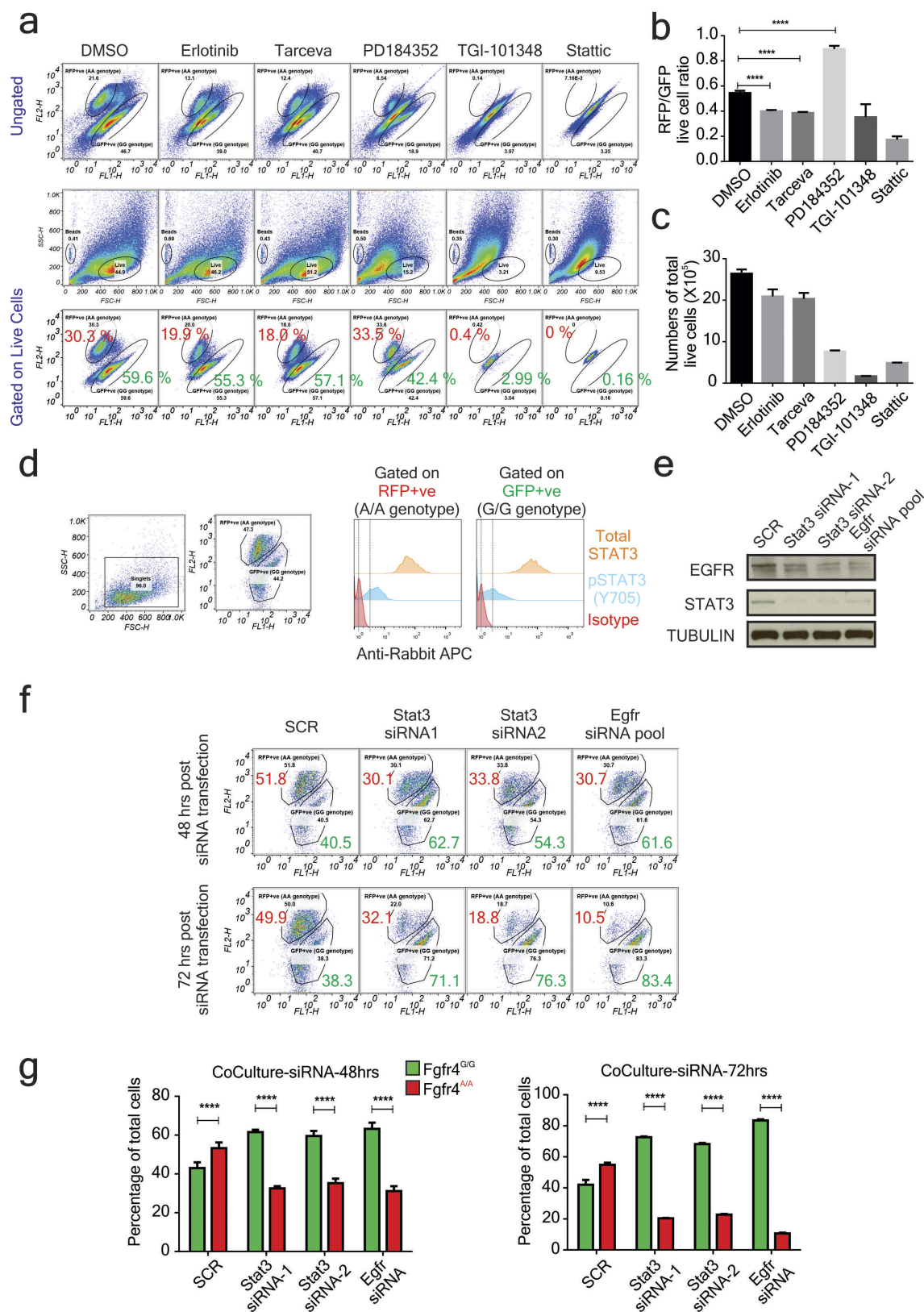
BiP/GRP78: red; nucleus: blue ($n = 12$, *Fgfr4*^{G/G} cells and $n = 15$, *Fgfr4*^{A/A} cells). **f**, Co-localization rate (means \pm s.d., $n = 12-15$, *** $P < 0.001$). **g**, Relative abundance of isoleucine (top), glycine (middle) and arginine (bottom) along the TMDs in plants (left column), fungi (middle column) and vertebrates (right column). The position relative to the cytosolic edge (arrow) of the TMD is on the horizontal axes, and the residue abundance is on the vertical axes. Graphical plots generated using an algorithm available at <http://www.tmdsonline.org>. **h**, Prediction of putative transmembrane segment in FGFR4 p.388G (black) and risk-variant FGFR4 p.388R (red) variants. Data obtained using algorithm at <http://split4.pmfst.hr/split/4/>.



Extended Data Figure 4 | Identification of germline variants in CD molecules generating membrane-proximal STAT3 binding site.

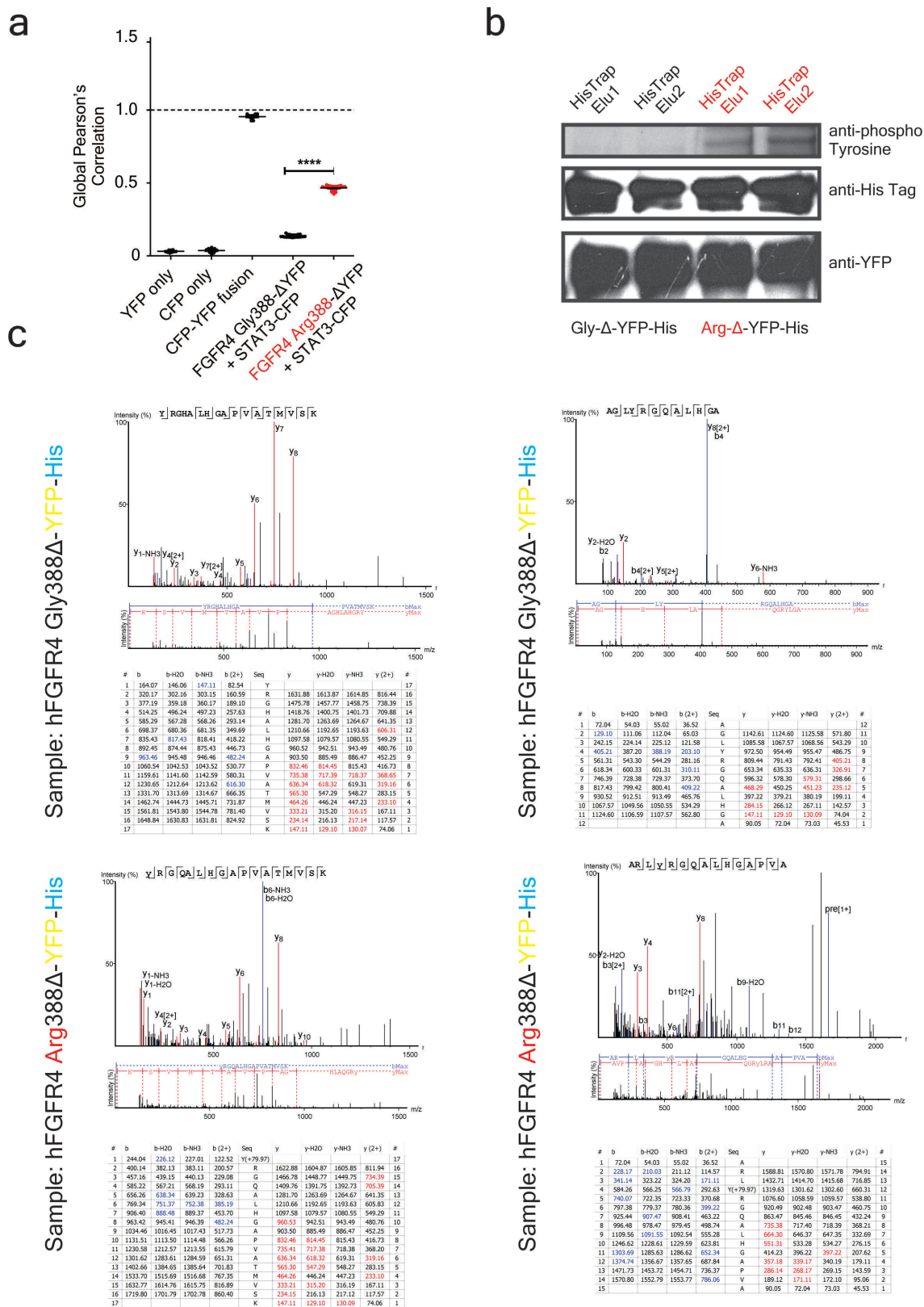
a, Graphical representation of the transmembrane sequence alignment of all human single-pass type I membrane proteins. Consensus sequence logo depicts stacks of amino-acid symbols with symbol height indicating the relative frequency of the amino acid in that position. Tyrosine (Y): green; glutamine (Q): magenta. **b**, Germline variations (superscript) in human CD molecules that generate a membrane-proximal STAT3 binding motif. Transmembrane domain: grey; missense mutation: red; YXXQ motifs: underlined. **c**, SNP name, protein allele, somatic status and the associated cancers for variants indicated by superscript numbers. **d**, STAT3-dependent promoter activity in immortalized *Fgfr4*^{G/G} and

Fgfr4^{A/A} MEFs stably expressing the indicated human oncogenes (means \pm s.d., $n = 6$, two-tailed non-parametric t -test, $*P < 0.05$, $**P < 0.01$ and $***P < 0.001$). **e**, Immunoblot analyses for expression of pSTAT3 (Y705) in immortalized *Fgfr4*^{G/G}, *Fgfr4*^{G/A} and *Fgfr4*^{A/A} MEFs stably expressing the indicated human oncogenes. **f**, Immunoblot analyses for expression of FGFR4 after knockdown of STAT3 in *Fgfr4*^{G/G}, *Fgfr4*^{G/A} and *Fgfr4*^{A/A} MEFs. **g**, Proliferation analyses of *Fgfr4*^{G/G}, *Fgfr4*^{G/A} and *Fgfr4*^{A/A} MEFs after STAT3 knockdown (means \pm s.e.m., $n = 5$, $*P < 0.05$, $**P < 0.01$ and $***P < 0.001$, two-way ANOVA and Tukey's multiple comparison test). Note the increased sensitivity of *Fgfr4*^{A/A} MEFs to the suppression of proliferation upon STAT3 knockdown.



Extended Data Figure 5 | Effect of rs351855 SNP on sensitivity to growth inhibitors. **a**, Flow cytometry analyses of co-cultivated *Fgfr4*^{G/G} (GFP⁺) and *Fgfr4*^{A/A} (RFP⁺) MEFs treated with the indicated growth-inhibiting drugs. Scatter plot depicts total cells (top) and live cells (bottom). **b**, Quantification of the relative proportion of live cells remaining after treatment (means \pm s.d., $n = 10$, two-tailed non-parametric t -test, *** $P < 0.001$). **c**, Quantification of total *Fgfr4*^{G/G} (GFP⁺) and *Fgfr4*^{A/A} (RFP⁺) MEF live cells remaining in the co-culture, indicative of the success of growth inhibition. **d**, Expression of pSTAT3 (Y705)

and total STAT3 expression in the co-cultivated *Fgfr4*^{G/G} (GFP⁺) and *Fgfr4*^{A/A} (RFP⁺) MEFs. **e**, Knockdown of STAT3 and EGFR expression in co-cultivated MEFs. **f**, Flow cytometry analyses of co-cultivated *Fgfr4*^{G/G} (GFP⁺) and *Fgfr4*^{A/A} (RFP⁺) MEFs after siRNA transfection. Dot-plot images are representative of three independent experiments. **g**, Relative quantification of *Fgfr4*^{G/G} (GFP⁺) and *Fgfr4*^{A/A} (RFP⁺) MEFs (means \pm s.d., $n = 6$, two-way ANOVA and Sidak's multiple comparison test, *** $P < 0.0001$).

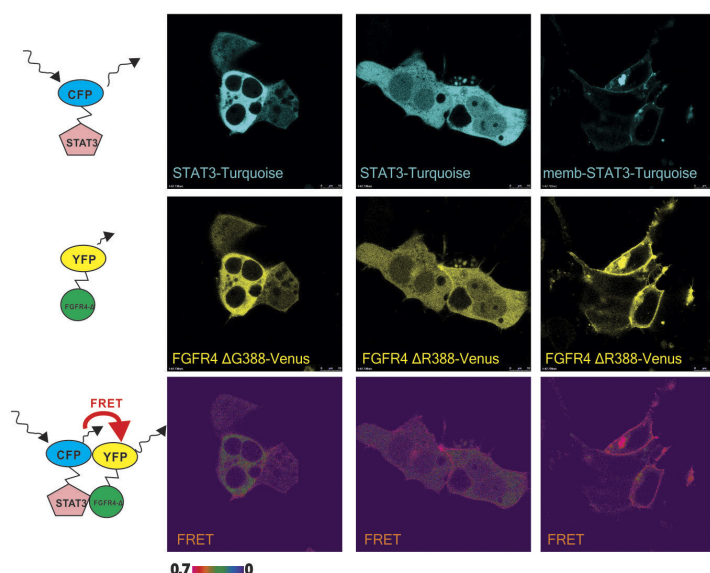


Extended Data Figure 6 | Identification of tyrosine-390 phosphorylation in the FGFR4 Arg388 variant. **a**, Co-localization analyses for STAT3-CFP and FGFR4 Gly388Δ-YFP and FGFR4 Arg388Δ-YFP variants in HEK293T transfectants (means \pm s.e.m., $n = 26$, **** $P < 0.0001$, one-tailed unpaired t -test with Welch's correction).

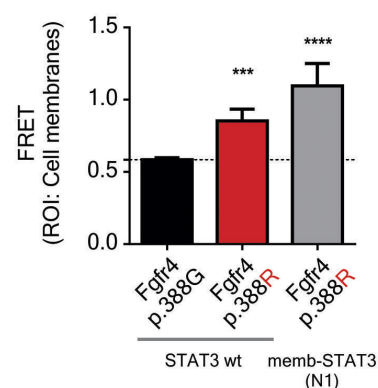
b, Immunoblot detection of phosphorylated tyrosines in purified FGFR4

Gly388Δ-YFP-His and FGFR4 Arg388Δ-YFP-His recombinant proteins. **c**, Mass spectrometry identification of Y-390 in FGFR4 Arg388Δ-YFP variant. Selected peptide-spectrum matches and the ion table displaying the calculated mass of the possible fragment ions are shown. N-terminal ions: blue; C-terminal ions: red.

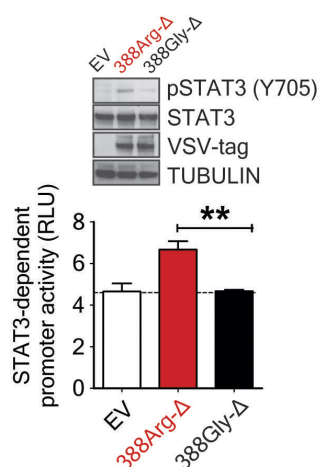
a



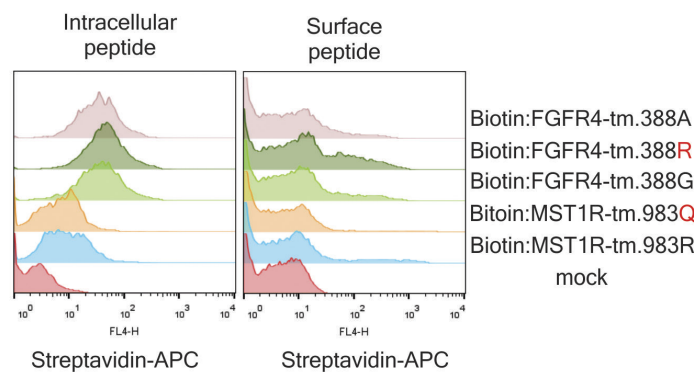
b



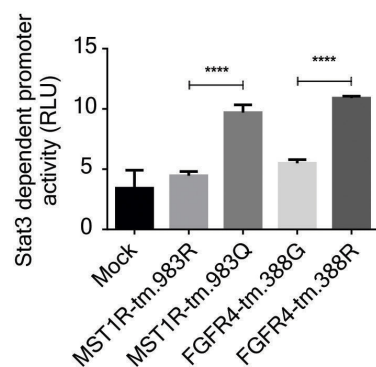
c



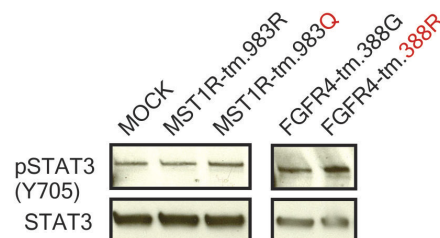
d



e

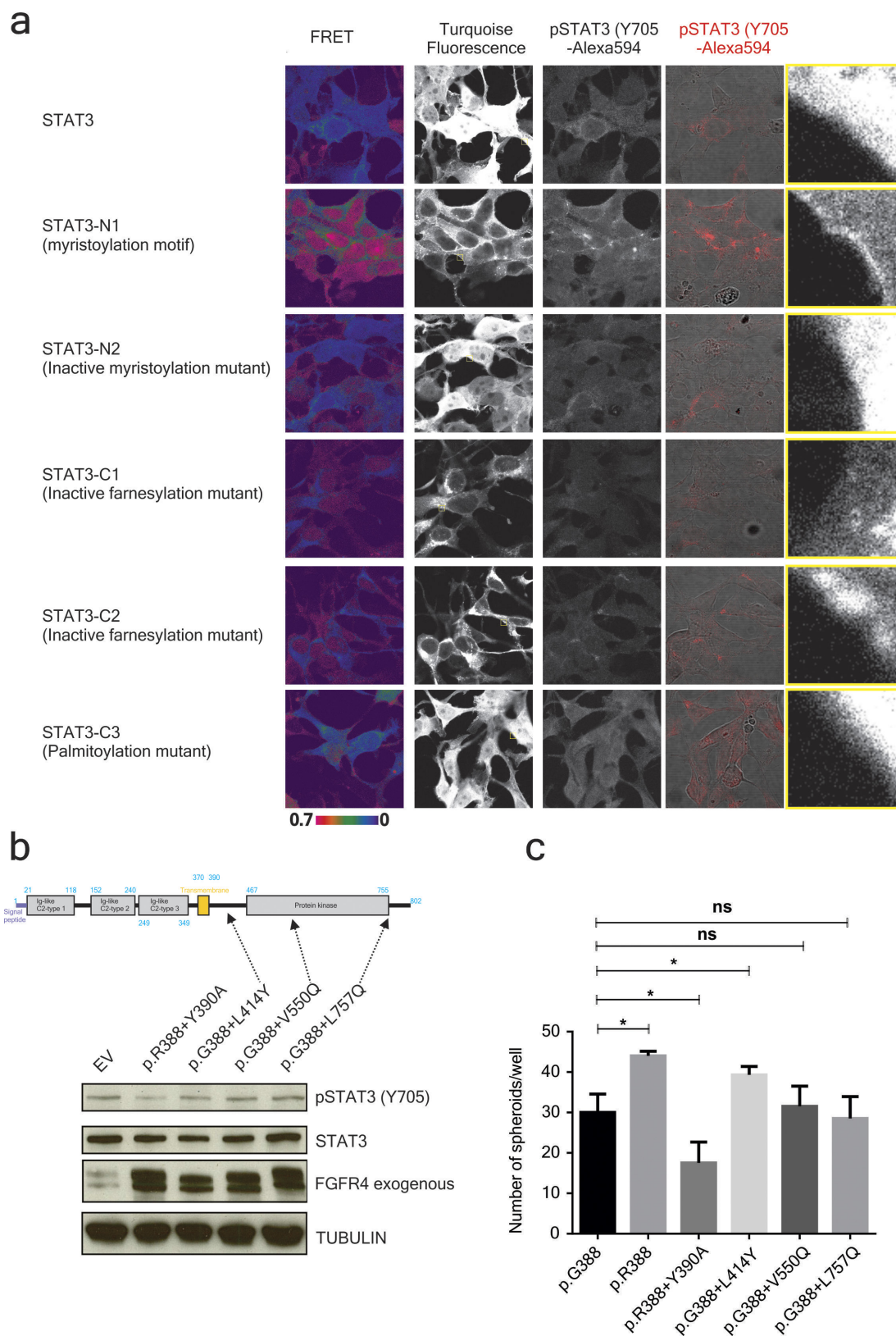


f



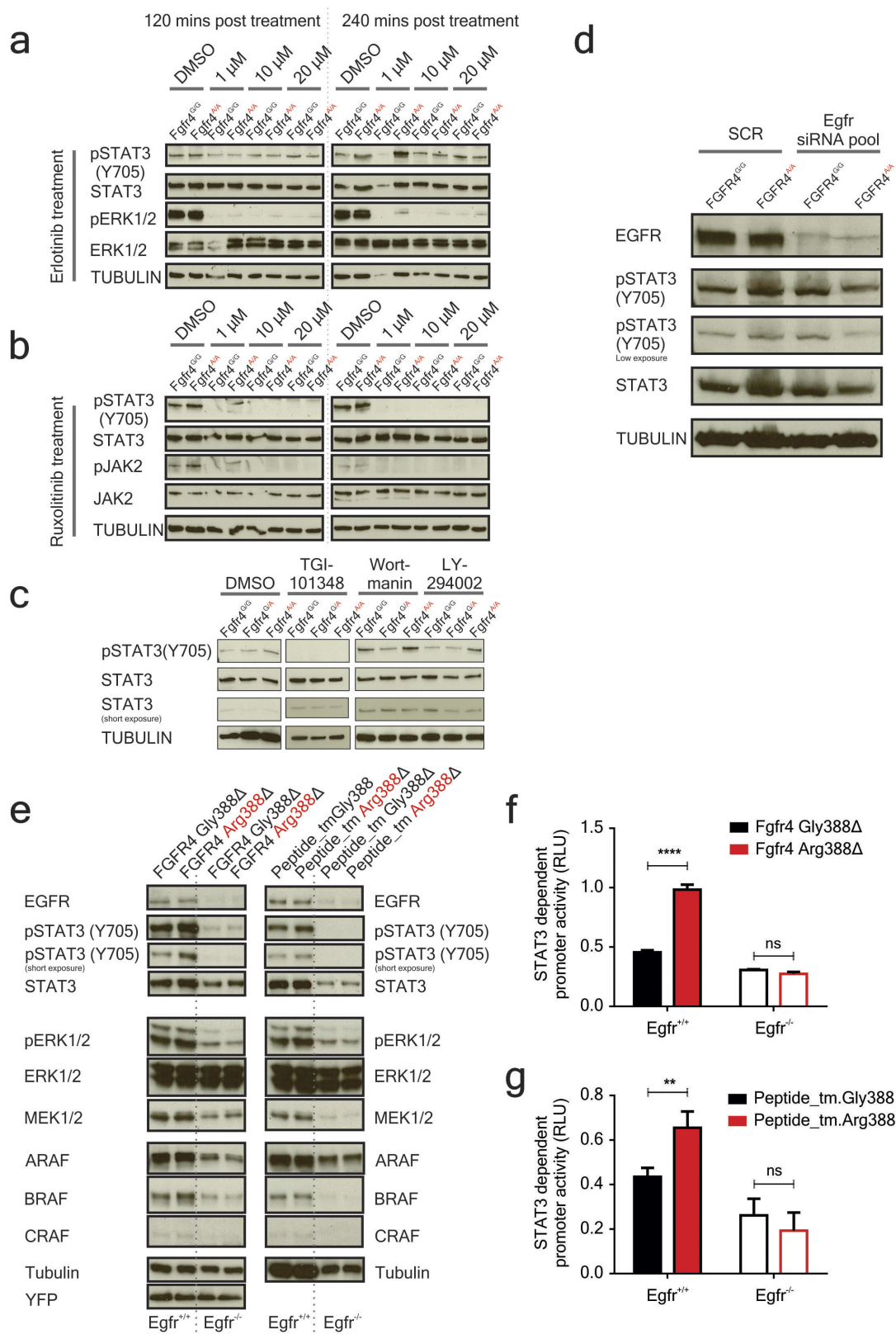
Extended Data Figure 7 | Direct interaction of STAT3 with the membrane-proximal YXXQ motifs. **a**, Representative CFP, YFP and FRET ratio images of HEK293T transfectants. Co-transfection with FGFR4 Arg388Δ-YFP and STAT3-turquoise-N1 (membrane-targeted STAT3) served as a reference control for the FRET calculations. Data shown are representative of five independent FRET imaging experiments. **b**, Quantification of FRET efficiencies calculated for the selected cell membrane as region of interest (ROI) (see Methods). Data shown are representative of five independent FRET imaging experiments

(means \pm s.e.m., $n = 12$ cells; *** $P < 0.001$, **** $P < 0.0001$). **c**, Quantification of STAT3-dependent promoter activity by FGFR4 Arg388 variant lacking intracellular kinase domain in HEK293T transfectants (means \pm s.e.m., $n = 6$, ** $P < 0.01$, two-tailed unpaired t -test, 95% confidence level). **d**, Assessment of intracellular and surface levels of the transfected transmembrane peptides by flow cytometry (see Methods). **e**, STAT3-dependent promoter activity in HEK293T cells transfected with the indicated peptides (means \pm s.e.m., $n = 3$, *** $P < 0.001$). **f**, Immunoblot analyses for pSTAT3 (Y705) expression in peptide transfectants.



Extended Data Figure 8 | Localization and phosphorylation of membrane-targeted STAT3. **a**, Representative confocal images of HEK293T transfectants. First column: FRET ratiometric images; second column: STAT3 localization; third column: pSTAT3 (Y705) localization; fourth column: overlay of pSTAT3 (Y705) and differential interference contrast images; fifth column enlarged images from selected (yellow rectangle) region. Magnification $\times 63$. Images are representative of ten acquisitions. **b**, Immunoblot analyses of pSTAT3 (Y705) expression

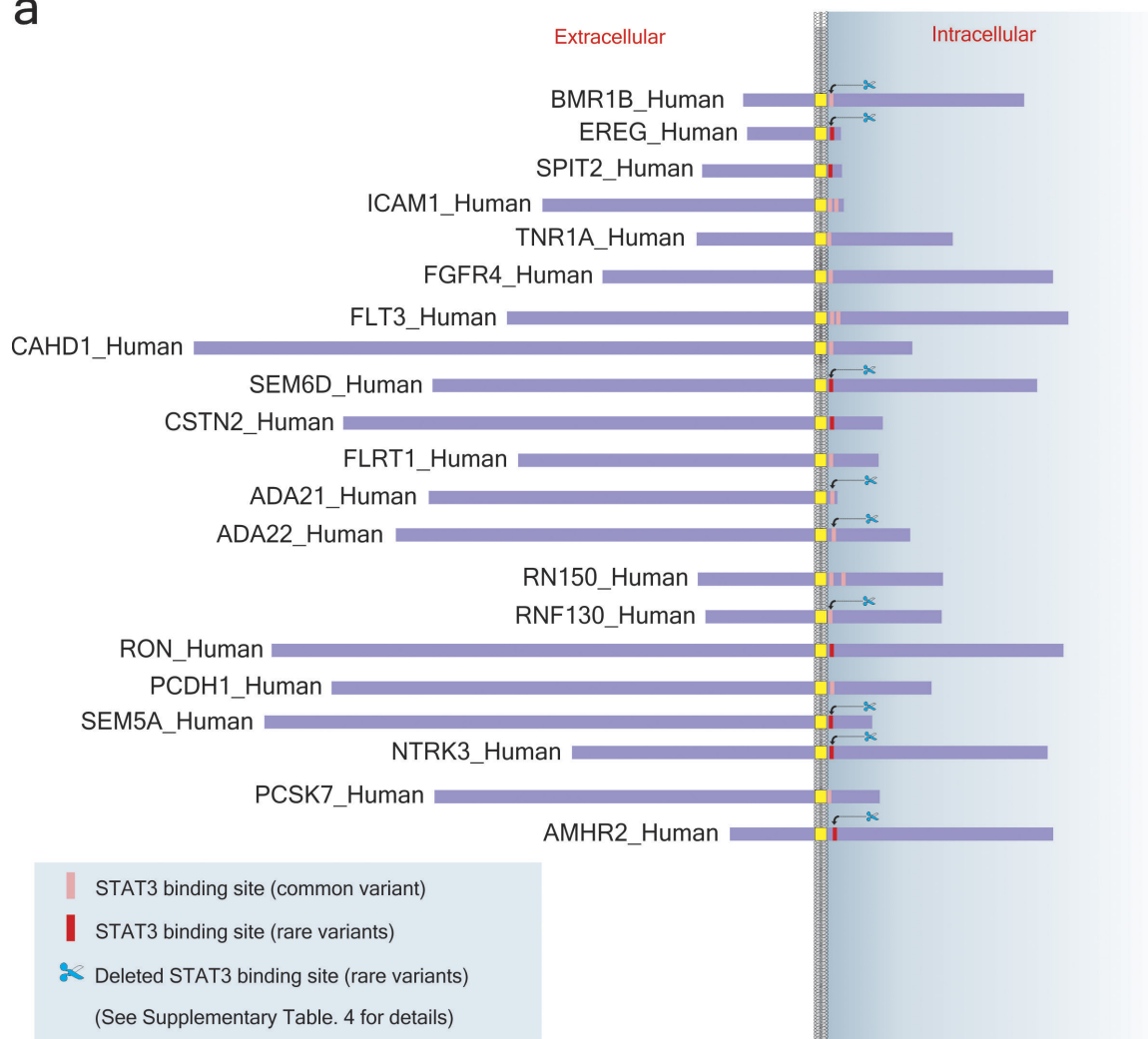
in HEK293T transfectants. The YXXQ motif was introduced at the juxtamembrane region (L414Y), tyrosine kinase domain (V550Q) and cytoplasmic tail terminus (L757Q) in the wild-type FGFR4 p.G388 variant. As a control, the YXXQ motif in risk variant FGFR4 p.R388 was destroyed by a mutation of Y390A. **c**, Quantification of colony formation assay results (see Methods). Shown are the representative results of three independent assays (means \pm s.e.m., $n = 4$ wells, $*P < 0.05$; ns, not significant).



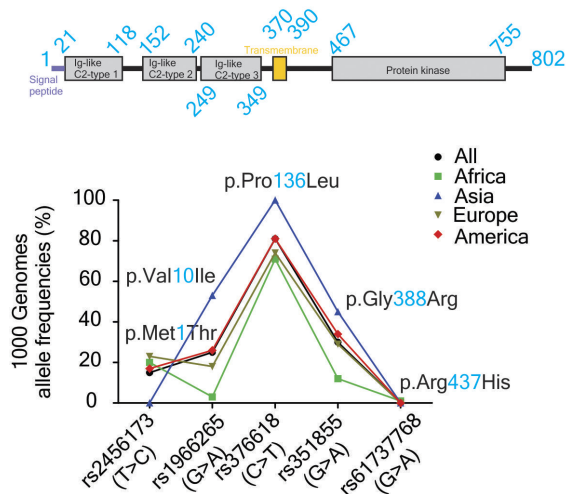
Extended Data Figure 9 | FGFR4 p.G388R-induced STAT3 activation is dependent on EGFR. **a–c**, Immunoblot analyses for pSTAT3 (Y705) expression in *Fgfr4*^{G/G} and *Fgfr4*^{A/A} MEFs treated with EGFR inhibitor (**a**), JAK inhibitor (**b**) and 10 μ M TGI-101348 (JAK inhibitor), 1 μ M wortmannin and 50 μ M LY24002 (PI3K inhibitor) (**c**). **d**, Immunoblot analyses for pSTAT3 (Y705) levels in *Fgfr4*^{G/G} and *Fgfr4*^{A/A} MEFs after EGFR knockdown. **e**, Immunoblot analyses for pSTAT3 (Y705) levels in

Egfr^{+/+} and *Egfr*^{-/-} MEFs transfected either with plasmids encoding the *Fgfr4* Gly388 Δ -YFP and *Fgfr4* Arg388 Δ -YFP variants or with the transmembrane peptides, namely peptide-tmGly388 (rs351855-G) and peptide-tmArg388 (rs351855-A). **f**, **g**, STAT3-dependent promoter activity in *Egfr*^{+/+} and *Egfr*^{-/-} MEFs transfected with plasmids (**f**) and peptides (**g**). The results are representative of three independent experiments (means \pm s.e.m., $n = 3$, *** $P < 0.001$ and ** $P < 0.01$).

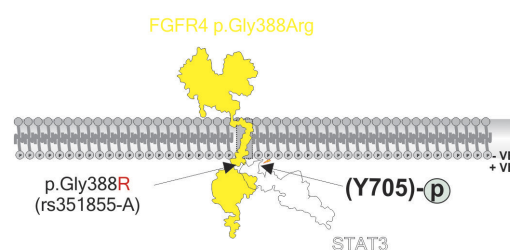
a



b



c



Extended Data Figure 10 | Human germline variants affecting the membrane-proximal STAT3 binding site. **a**, Summary of the results obtained from combined analyses of the Ensembl variation data set and NHLBI exome variant data set. Dim red boxes: common germline mutations; bright red boxes: rare germline mutations that give rise to a YXXQ motif. Scissored arrowheads: rare germline mutations that destroy the YXXQ motif either by a frame shift or deletion at or before the YXXQ motif in the DNA sequence. **b**, 1000 Genome allele frequencies of all the *FGFR4* non-synonymous coding region germline variants. We used

the data from the 1000 Genomes Project (released 14 October 2013). **c**, Graphical summary explaining the new molecular function acquired by the germline variant rs351855 in the *FGFR4* transmembrane domain. Alteration of the *FGFR4* transmembrane spanning segment, such that Y³⁹⁰ was now located in the cytoplasm and phosphorylated, thereby exposing the functional STAT3 binding site (Y³⁹⁰-[p]RGQ³⁹⁰). Consequently, membrane-proximal phosphate transfer reactions (yellow symbol) dependent on EGFR activity lead to STAT3 tyrosine-705 phosphorylation, resulting in enhanced STAT3 signalling in cells.

Competition between DNA methylation and transcription factors determines binding of NRF1

Silvia Domcke^{1,2*}, Anaïs Flore Bardet^{1*}, Paul Adrian Ginno¹, Dominik Hartl^{1,2}, Lukas Burger^{1,3} & Dirk Schübeler^{1,2}

Eukaryotic transcription factors (TFs) are key determinants of gene activity, yet they bind only a fraction of their corresponding DNA sequence motifs in any given cell type¹. Chromatin has the potential to restrict accessibility of binding sites; however, in which context chromatin states are instructive for TF binding remains mainly unknown^{1,2}. To explore the contribution of DNA methylation to constrained TF binding, we mapped DNase-I-hypersensitive sites in murine stem cells in the presence and absence of DNA methylation. Methylation-restricted sites are enriched for TF motifs containing CpGs, especially for those of NRF1. In fact, the TF NRF1 occupies several thousand additional sites in the unmethylated genome, resulting in increased transcription. Restoring *de novo* methyltransferase activity initiates remethylation at these sites and outcompetes NRF1 binding. This suggests that binding of DNA-methylation-sensitive TFs relies on additional determinants to induce local hypomethylation. In support of this model, removal of neighbouring motifs in *cis* or of a TF in *trans* causes local hypermethylation and subsequent loss of NRF1 binding. This competition between DNA methylation and TFs *in vivo* reveals a case of cooperativity between TFs that acts indirectly via DNA methylation. Methylation removal by methylation-insensitive factors enables occupancy of methylation-sensitive factors, a principle that rationalizes hypomethylation of regulatory regions.

Methylation of DNA at cytosines within CpG dinucleotides has the potential to block TF binding either directly through interference with base recognition or indirectly through recruitment of methylation-specific binding proteins³. DNA methylation has been reported to block binding of some TFs *in vitro*³. However, this does not necessarily translate to a similar effect *in vivo*^{4,5}. In addition, sensitivity *in vivo* can be highly locus-specific as observed for the TF CTCF, which only responds to methylation at a very limited set of chromosomal loci^{6–9}. Intriguingly, some TFs such as REST and CTCF have been shown to bind methylated regions and trigger their demethylation^{8,10,11}. Thus, although it is established that active regulatory regions are bound by TFs and generally display low levels of DNA methylation^{8,12}, it remains contentious whether this relationship reflects the cause or consequence of altered TF binding^{13,14}. Determining factor-specific sensitivity of binding events in a cellular context is therefore imperative for understanding how DNA methylation affects gene expression and to functionally interpret epigenomic maps. To identify TFs that are restricted in their binding by DNA methylation *in vivo*, we mapped DNase-I-hypersensitive sites (DHSs), an indicator of TF binding, in wild-type murine embryonic stem (ES) cells and upon global removal of DNA methylation (Fig. 1a).

DNA methylation is essential for mouse development and survival of most tested mammalian cell types, with the exception of murine ES cells¹⁵. Therefore, these cells provide an opportunity to compare TF binding in the presence and absence of DNA methylation. To reduce genetic or clonal variability we used CRISPR/Cas9 to generate genetic deletions of both *de novo* DNA methyltransferases

Dnmt3a and *Dnmt3b* and the maintenance enzyme *Dnmt1* in the ES cell line 159 (see Methods) for which we previously performed base-pair-resolution methylation profiling⁸ (Extended Data Fig. 1a).

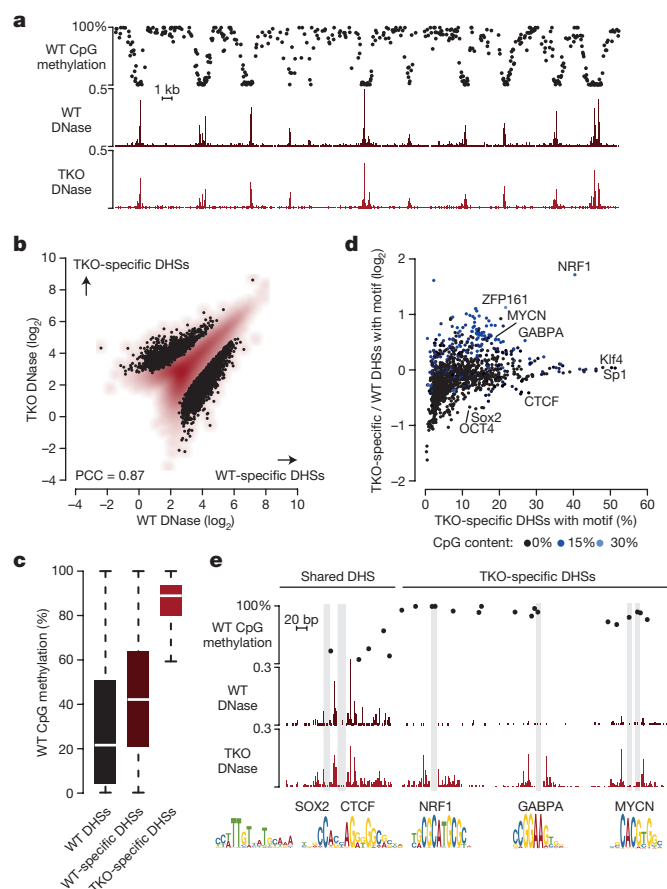


Figure 1 | DHSs that form upon removal of DNA methylation are enriched for specific TF motifs. **a**, Wild-type (WT) methylation, and wild-type and TKO DNase-seq signal at a representative genomic region (chr17: 25,920,000–25,972,499). **b**, DNase-seq signal at all DHSs in wild type and TKO. Black dots mark DHSs significantly enriched in wild type ($n = 2,837$) or TKO ($n = 1,543$). PCC, Pearson correlation coefficient. **c**, Average wild-type methylation of CpGs within all wild-type, wild-type-specific (the subset of wild-type DHSs that are not present in TKO DHSs) or TKO-specific DHSs. Boxplots show median (white line), 25th and 75th percentiles (boundaries), minimum and maximum (whiskers). **d**, Motif occurrences in TKO-specific DHSs compared to all wild-type DHSs. Blue colouring illustrates motif CpG content. **e**, Representative genomic regions showing shared (left, chr6: 31,189,871–31,190,470) and TKO-specific (right, chr1: 51,483,272–51,483,871, chr6: 48,413,300–48,413,899 and chr10: 62,623,300–62,623,899) DHS footprints. Motif locations are highlighted in grey.

¹Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH 4058 Basel, Switzerland. ²University of Basel, Faculty of Sciences, Petersplatz 1, CH 4003 Basel, Switzerland. ³Swiss Institute of Bioinformatics, Maulbeerstrasse 66, CH 4058 Basel, Switzerland.

*These authors contributed equally to this work.

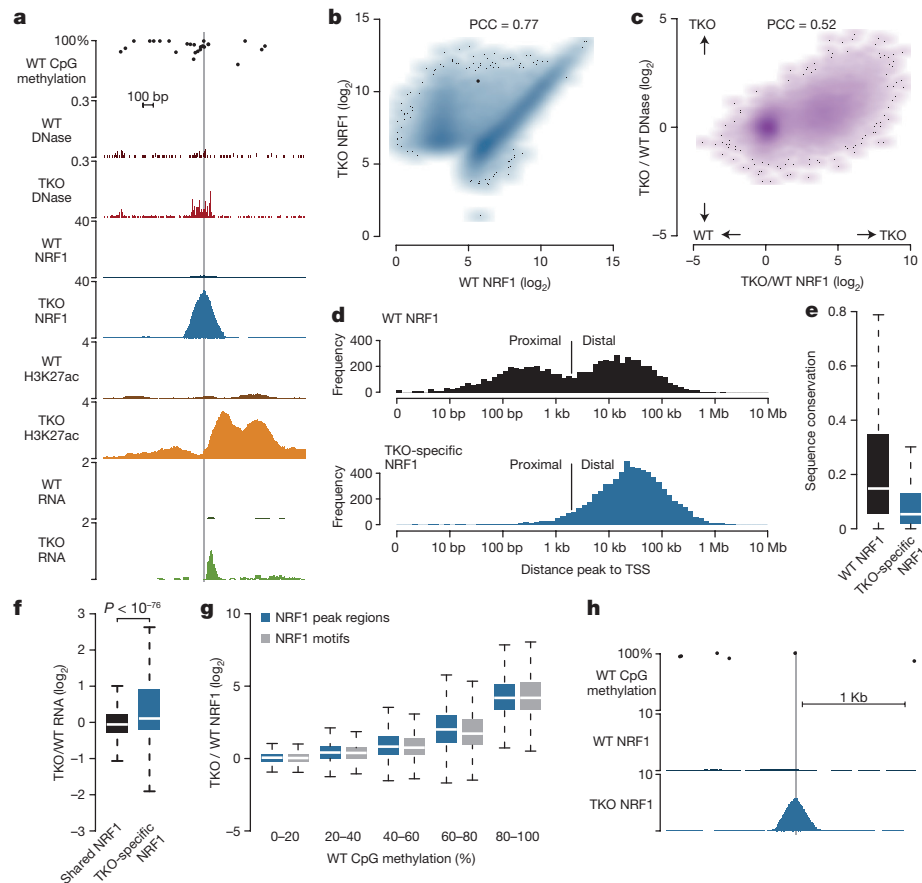


Figure 2 | NRF1 binds several thousand new sites in the unmethylated genome. **a**, Wild-type methylation, and wild-type and TKO DNase-seq, NRF1 ChIP-seq, H3K27ac ChIP-seq and RNA-seq signal at a TKO-specific distal genomic region (chr4: 99,235,857–99,236,456). The NRF1 motif location is highlighted in grey. **b**, Wild-type and TKO NRF1 ChIP-seq signal at all peak regions. The thick black dot represents the region in **a**. **c**, Changes in NRF1 binding and DNase-seq signal between wild type and TKO at all NRF1 peak regions. **d**, Distance of all wild-type (top; $n = 8,835$) or TKO-specific (bottom; $n = 7,205$) NRF1 peaks to the nearest transcriptional start site (TSS). Cutoff between proximal and distal sites is 2 kb. **e**, Average sequence conservation (PhastCons score) of all

wild-type or TKO-specific NRF1 peak regions. Boxplots show median (white line), 25th and 75th percentiles (boundaries), minimum and maximum (whiskers). **f**, Expression change (in reads per kilobase per million (RPKM)) of genes closest to shared and TKO-specific NRF1 peaks. P value from a Wilcoxon test. **g**, Change in NRF1 binding between TKO and wild-type at all peak regions grouped according to their average methylation. Blue boxes represent changes within entire peak regions, grey boxes only those within NRF1 motifs. $n > 800$ in all groups. **h**, Wild-type methylation, and wild-type and TKO NRF1 ChIP-seq signal at a genomic region with no additional CpGs within 1.8 kb around the motif (grey line).

The resulting triple knockout (TKO) cells showed no detectable DNA methylation by several measures (Extended Data Fig. 1b, c) and limited changes in global expression patterns, as previously reported for a TKO cell line generated by classical mouse genetics^{15,16} (Extended Data Fig. 1d, e).

Hypersensitivity to digestion by DNase I is an indicator of TF binding that does not require a priori knowledge of the TFs involved¹⁷. We mapped DHSs with high coverage in both wild-type cells and the isogenic TKO cells and observed that the vast majority of DHSs remain unchanged (Fig. 1a, b, Extended Data Fig. 2a–d and Extended Data Table 1). This suggests that the binding patterns of most TFs expressed in murine ES cells are not altered upon global removal of DNA methylation. In addition, we observed a fraction of DHSs that are specific to each cell state in a reproducible manner (Fig. 1b and Extended Data Fig. 2e). These DHSs are preferentially located distal to transcriptional start sites (TSS) and within CpG-poor regions (Extended Data Fig. 2f, g). In contrast to wild-type-specific DHSs (the subset of wild-type DHSs that are not present in TKO DHSs), newly formed sites in the TKO cell line lie within regions that were methylated in the wild-type cells, indicating that they could be methylation-dependent (Fig. 1c and Extended Data Fig. 2h).

Searching for known TF motifs and hexamer sequences enriched in TKO-specific DHSs resulted in a small number of candidate

methylation-sensitive TFs including NRF1, GABPA and MYCN (Fig. 1d, Extended Data Fig. 3a and Supplementary Table 1). These factors are expressed at similar levels in both cell lines and probably form TKO-specific DNase I footprints (Fig. 1d, e and Extended Data Fig. 3b, c). In contrast, motifs enriched in wild-type-specific DHSs do not reveal footprints limited to this cell state (Fig. 1b and Extended Data Fig. 3c, d). Notably, TKO-specific DHSs are enriched for motifs containing CpG dinucleotides, even though they reside within regions that are generally CpG-poor (Fig. 1d, Extended Data Fig. 2g and Supplementary Table 1). The most prominently enriched motif in TKO-specific DHSs contains two CpGs, consistent with a direct inhibition by DNA methylation, and belongs to the highly conserved TF nuclear respiratory factor 1 (NRF1)¹⁸ (Fig. 1d, e). Previous *in vitro* experiments with NRF1 suggested that DNA methylation blocks binding^{19,20}, but also that it preferentially binds to methylated sequences²¹. Given its strong signal and because only one factor has been reported to bind this motif, we focused on further analysis of NRF1.

Chromatin immunoprecipitation of NRF1 followed by sequencing (ChIP-seq) revealed that more than 7,000 sites, in addition to those already occupied in wild-type cells, show reproducible increased NRF1 binding in the absence of DNA methylation (Fig. 2a, b, Extended Data Fig. 4a–d and Extended Data Table 1). Newly bound NRF1 sites correlate with TKO-specific DHSs, validating the comparative DHS

approach (Fig. 2a, c). They occur distal to genes (Fig. 2d) in regions of low CpG content (Extended Data Fig. 4e) and poor sequence conservation (Fig. 2e), suggesting that a large fraction could represent non-functional sites otherwise blocked by DNA methylation. Nevertheless, increase of NRF1 binding is matched by a significant increase in expression of the nearest genes, indicative of an impact on transcription (Fig. 2f). Additionally, for some TKO-specific sites, lysine 27 acetylation of histone H3, a mark of active regulatory regions, appears and aberrant NRF1-dependent transcripts are initiated directly at the binding sites (Fig. 2a and Extended Data Fig. 4a, f–j).

TKO-specific NRF1 sites mostly contain a high confidence motif with at least one but usually two CpGs (Extended Data Fig. 4k and Supplementary Table 2). These motifs display intermediate to full methylation in wild-type cells, yet increased NRF1 binding in TKO is strongest at highly methylated motifs, suggesting that methylation of the core motif directly prevents binding in wild-type cells (Fig. 2g and Extended Data Fig. 4l). TKO-specific binding of NRF1 is independent of the density of methylated CpGs in the surrounding region, strongly arguing against an involvement of indirect repression through methyl-CpG binding-domain proteins²² (Extended Data Fig. 4m–o). This is exemplified at a locus that harbours no CpG within 1.8 kb around the motif (Fig. 2h); despite this absence of additional CpGs, NRF1 binds in a strictly methylation-dependent manner.

In the experiments described so far, ES cells were cultured in the presence of serum and LIF, which recapitulates the genome-wide methylation observed in the postimplantation epiblast²³. Culturing in the presence of two kinase inhibitors (2i) is an alternative regime that mimics the inner cell mass of the blastocyst and coincides with downregulation of the *de novo* *Dnmts*^{24,25}. Here it provides the opportunity to measure NRF1 binding at physiological levels of low methylation and without genetic alteration of the *Dnmt* genes. Transferring wild-type cells cultured originally in serum to 2i conditions leads to increased NRF1 binding at the vast majority of previously identified TKO-specific sites (Fig. 3a, b and Extended Data Fig. 5a–c). Similarly, this coincides with hypomethylation of these sites in 2i conditions as revealed by whole-genome, as well as high-coverage amplicon, methylation profiling (Fig. 3a and Extended Data Fig. 5d–f). Small differences in NRF1 binding between 2i and TKO conditions are readily explained by remaining levels of methylation at a subset of sites in 2i (Extended Data Fig. 5c, g). These include examples where the motif remains methylated and unbound even though the surrounding region is demethylated (Extended Data Fig. 5h), providing additional support for our observation that methylation of the core motif alone is the critical determinant of NRF1 binding *in vivo*.

To test if NRF1 binding to these new sites inhibits their *de novo* methylation, we transferred ES cells cultured in 2i back to medium with serum. This leads to transcriptional upregulation of the *de novo* *Dnmt* genes and genomic remethylation over time²⁵. Profiling of NRF1 binding, as well as whole-genome and amplicon methylation, revealed that the majority of methylation-dependent sites become remethylated and that NRF1 binding can no longer be detected (Fig. 3a, c and Extended Data Fig. 5h–m). This shows that *de novo* methylation can outcompete binding of NRF1, implying that binding and creation of a DHS is not sufficient to protect against *de novo* methylation for this TF.

Although levels of *Nrf1* expression remained mostly unchanged between the tested conditions (Extended Data Fig. 3b and Extended Data Fig. 5a), we assessed if variations in NRF1 protein abundance could account for differential occupancy. Therefore we overexpressed NRF1 at least tenfold and profiled its genomic binding in wild-type cells (Extended Data Fig. 6a). This revealed an increase in binding at previously occupied sites but also novel sites (Fig. 3a and Extended Data Fig. 6b, c). The latter, however, do not overlap with methylation-dependent sites and contain weak NRF1 motifs, reflecting less specific binding to regions of open chromatin (Fig. 3a and Extended Data Fig. 6d, e). This shows that methylation of individual core motifs, and not NRF1 protein levels, determines genomic occupancy.

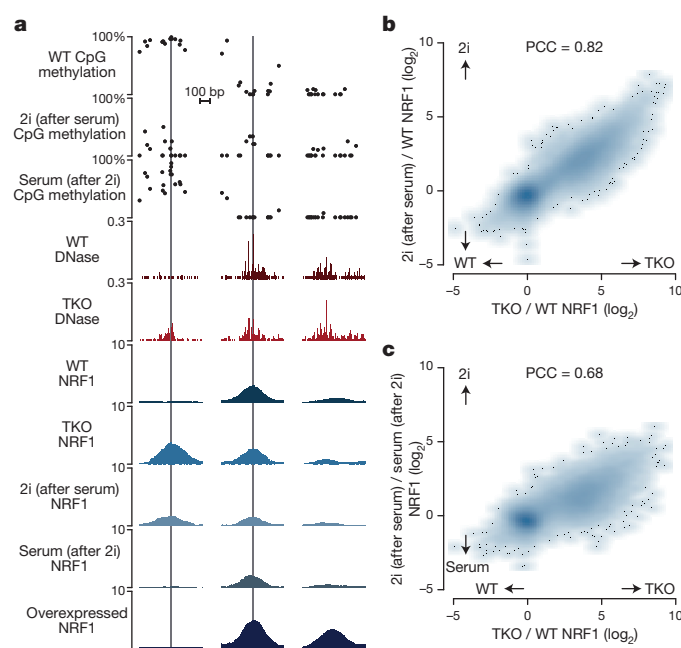


Figure 3 | *De novo* methylation outcompetes NRF1 binding. **a**, Wild-type methylation, wild-type and TKO DNase-seq, and NRF1 ChIP-seq signal in wild-type under different culture regimes, in TKO and in wild-type cells overexpressing NRF1 at representative genomic regions. Left, TKO-specific site (chr12: 82,788,342–82,794,341). Middle, shared site in wild-type and TKO, with increased binding upon overexpression (chr5: 148,104,611–148,105,210). Right, site only bound upon overexpression (chr18: 36,030,688–36,036,687). Grey lines indicate the location of the NRF1 motif. **b**, Change of NRF1 ChIP-seq signal between wild-type and TKO or 2i culture (after culture with serum) at all NRF1 peaks regions. **c**, Change of NRF1 ChIP-seq signal between TKO and wild-type versus between culture with 2i (after culture with serum) and culture with serum (after culture with 2i) at all NRF1 peak regions.

To test if cell-type-specific methylation patterns could similarly explain differential binding of NRF1, we differentiated ES cells into neuronal progenitors and investigated NRF1 binding. We found that the gain of methylation at NRF1 motifs in neuronal progenitors coincides with loss of NRF1 binding (Extended Data Fig. 7a–c) and matching lower expression of neighbouring genes (Extended Data Fig. 7d, e). This tight link between DNA methylation, NRF1 binding and transcription holds true beyond the murine system, as seen by genomic profiling of NRF1 in human normal breast cells (HMEC) and a breast cancer cell line (HCC1954)²⁶ (Extended Data Fig. 7f–i), as well as in other cell type comparisons (H1hESC and GM12878)²⁷ (Extended Data Fig. 7j–m). Thus, data from different organisms and cellular states including cancer may indicate that methylation-dependent binding of NRF1 is a general phenomenon that affects gene regulation.

We next sought to test the methylation sensitivity of NRF1 without global reduction of DNA methylation, by using reporter constructs inserted into a defined chromosomal locus of ES cells by Cre recombinase²⁸. NRF1 sites with 400 bp of their surrounding genomic sequence were inserted either unmethylated or premethylated at CpGs *in vitro* (Extended Data Fig. 8a–c). As expected, this revealed reduced binding to the premethylated compared to the untreated template (Extended Data Fig. 8b). Thus, sensitivity of NRF1 to methylation of the underlying motif can be recapitulated in an ectopic site. We previously showed that CTCF can bind a motif added to a premethylated reporter and cause local reduction of methylation⁸. When we exchanged the CTCF motif with that of NRF1, we did not observe NRF1 binding or loss of methylation. Only upon forced demethylation is NRF1 capable of binding this minimal sequence context (Fig. 4a). Therefore NRF1 can bind its motif autonomously, but only if unmethylated. Genome-wide binding and single-locus reporter experiments indicate that NRF1 is

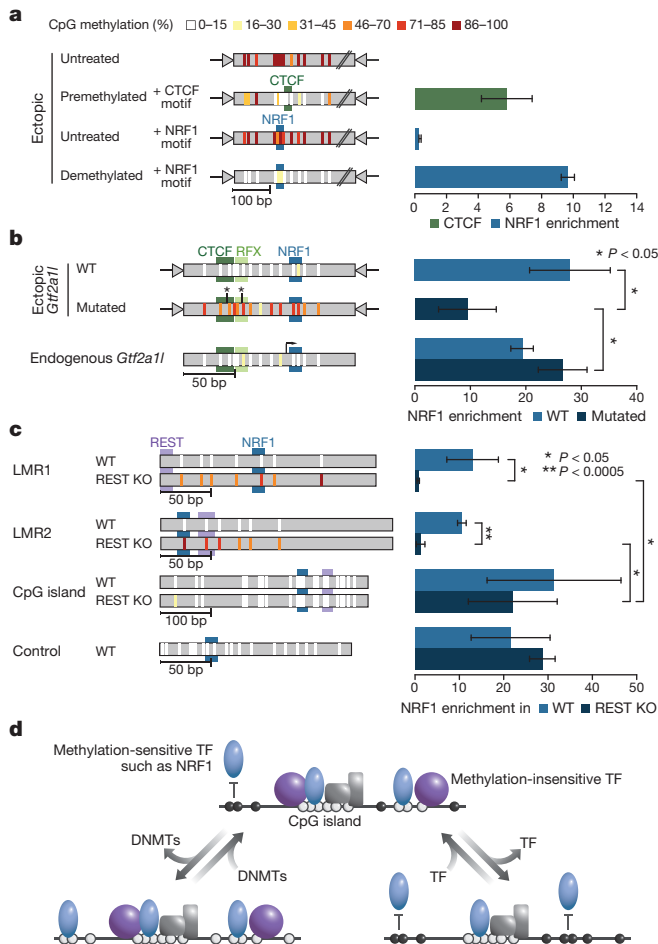


Figure 4 | NRF1 binds to unmethylated core motifs via TF-mediated local hypomethylation. **a–c**, Methylation levels of individual CpGs (left, amplicon Bis-seq) and TF occupancy (right, ChIP-qPCR) for reporters inserted into a defined ectopic genomic locus or for endogenous regions. TF motif locations are marked as coloured boxes. ChIP-qPCR enrichments are the mean of three biological replicates; error bars represent standard deviation; *P* values from two-sided *t*-tests. **a**, CTCF or NRF1 motifs were added to the same sequence inserted as either premethylated (CTCF)⁸, untreated or chemically demethylated (NRF1). **b**, The *Gtf2a1* promoter was inserted with intact or mutated (asterisks) CTCF and RFX motifs²⁸. In both cases the corresponding endogenous locus serves as control. **c**, Endogenous regions bound by REST in wild-type and containing adjacent NRF1 motifs in low-methylated regions (LMR) and an unmethylated CpG island, profiled in wild-type and REST knockout (KO) cells. The control region is REST independent. **d**, Image of the model. In wild-type cells, NRF1 binding is blocked by DNA methylation and only occurs at unmethylated motifs (top). Motif methylation requires the activity of the DNMTs (bottom left), while motif demethylation can be mediated upon adjacent binding of methylation-insensitive TFs (bottom right). Circles represent unmethylated (white) or methylated (black) CpGs.

sensitive to DNA methylation of its motif and that it cannot protect it from *de novo* methylation. This leads to the prediction that NRF1 relies on other features that keep its motif in an unmethylated state.

As some TFs, such as CTCF, can locally mediate low methylation levels^{8,28,29}, we hypothesized that such factors could direct NRF1 binding in wild-type cells. Consistent with this model, constitutive NRF1 binding sites reside in regions that are co-bound by many TFs, as reflected by broad DHSs and overlap with existing TF localization maps (Extended Data Fig. 9a, b). To experimentally test this hypothesis we inserted reporter constructs harbouring an endogenous promoter sequence including a NRF1 motif (Extended Data Fig. 9c). Deletion of the CTCF and RFX motifs within this construct leads to its hypermethylation²⁸ but notably also to decreased NRF1 binding (Fig. 4b).

This establishes a dependence of NRF1 *in cis* on motifs of TFs that mediate local hypomethylation. To further explore this hierarchical model, we assessed whether removal of a demethylating TF affects NRF1 binding. We previously showed that REST (also known as NSRF) creates regions of low methylation at its binding sites in CpG-poor regions, which become remethylated when REST is genetically removed^{8,10}. Even though REST and NRF1 have not been functionally linked, we identified a few sites where NRF1 binds adjacent to REST (Extended Data Fig. 9d), enabling us to monitor NRF1 occupancy as a function of REST. At sites that occur within CpG-poor low-methylated regions, we observe *de novo* methylation upon deletion of REST that extends well into the NRF1 motif and coincides with loss of NRF1 binding in both cases tested (Fig. 4c). Of note, the absence of REST does not affect proximal NRF1 binding within a CpG island, as it remains hypomethylated regardless of REST occupancy, possibly because CpG islands are bound by additional factors that confer hypomethylation²⁹ (Fig. 4c). Thus, NRF1 binding *in vivo* critically relies on the local DNA sequence context *in cis* and TFs *in trans* to ensure a hypomethylated binding site (Fig. 4d).

This study proposes several TFs that might be restricted by DNA methylation but also suggests that the majority of factors expressed in mouse ES cells do not respond to global loss of DNA methylation. A critical question remains whether differentiated cells, for which DNA methylation has been shown to be essential, express a larger set of methylation-sensitive factors.

Our study of NRF1 binding in different and dynamic methylomes establishes an example of genome-wide, methylation-sensitive TF binding *in vivo*. Combined with site-specific genetic and epigenetic perturbation, it provides a proof of principle for a model whereby DNA methylation can guide TF binding in a highly factor- and context-specific manner (Fig. 4d).

NRF1 has previously been proposed to be a pioneer factor based on its ability to form a DHS *de novo*³⁰. We show that NRF1 only bears canonical hallmarks of a pioneer factor² in the absence of DNA methylation, where it indeed can bind autonomously and form a DHS. In the presence of DNA methylation, it behaves as a ‘settler’ TF, as it requires the assistance of superordinate TFs to ensure hypomethylation of its motif. This suggests that the ability to mediate a hypomethylated state upon binding could be an additional relevant characteristic for a pioneer TF in vertebrates. Notably, we show that NRF1 binding to an unmethylated site does not protect against *de novo* methylation. This provides clear evidence for competition between TFs and DNMTs, and argues that active demethylation and/or efficient obstruction of *de novo* methylation is required not only for the establishment of NRF1 binding, but also for its maintenance. This exemplifies the idea that TF hierarchies can be mediated via a local epigenetic mark—DNA methylation removal by methylation-insensitive factors enables occupancy of methylation-sensitive factors in a form of indirect cooperativity that does not require physical interaction between both TFs¹. It illustrates that TF binding patterns at enhancers and promoters are both guided by and actively shape the balance between active demethylation and *de novo* methylation (Fig. 4d). This supports a model in which the role of DNA methylation in restricting genomic binding of TFs is dependent on the specific factor, the local activity of methylating and demethylating enzymes, and the genomic context of individual motif occurrences.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 April; accepted 16 November 2015.

Published online 16 December 2015.

- Slattery, M. *et al.* Absence of a simple code: how transcription factors read the genome. *Trends Biochem. Sci.* **39**, 381–399 (2014).
- Iwafuchi-Doi, M. & Zaret, K. S. Pioneer transcription factors in cell reprogramming. *Genes Dev.* **28**, 2679–2692 (2014).

3. Tate, P. H. & Bird, A. P. Effects of DNA methylation on DNA-binding proteins and gene expression. *Curr. Opin. Genet. Dev.* **3**, 226–231 (1993).
4. Becker, P. B., Ruppert, S. & Schütz, G. Genomic footprinting reveals cell type-specific DNA binding of ubiquitous factors. *Cell* **51**, 435–443 (1987).
5. Weih, F., Nitsch, D., Reik, A., Schütz, G. & Becker, P. B. Analysis of CpG methylation and genomic footprinting at the tyrosine aminotransferase gene: DNA methylation alone is not sufficient to prevent protein binding *in vivo*. *EMBO J.* **10**, 2559–2567 (1991).
6. Bell, A. C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* **405**, 482–485 (2000).
7. Hark, A. T. et al. CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. *Nature* **405**, 486–489 (2000).
8. Stadler, M. B. et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490–495 (2011).
9. Maurano, M. T. et al. Role of DNA methylation in modulating transcription factor occupancy. *Cell Rep.* **12**, 1184–1195 (2015).
10. Feldmann, A. et al. Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLoS Genet.* **9**, e1003994 (2013).
11. Wu, H. & Zhang, Y. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell* **156**, 45–68 (2014).
12. Ziller, M. J. et al. Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**, 477–481 (2013).
13. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Rev. Genet.* **13**, 484–492 (2012).
14. Schübeler, D. Function and information content of DNA methylation. *Nature* **517**, 321–326 (2015).
15. Tsumura, A. et al. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**, 805–814 (2006).
16. Karimi, M. M. et al. DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell Stem Cell* **8**, 676–687 (2011).
17. Neph, S. et al. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**, 83–90 (2012).
18. Virbasius, C. A., Virbasius, J. V. & Scarpulla, R. C. NRF-1, an activator involved in nuclear-mitochondrial interactions, utilizes a new DNA-binding domain conserved in a family of developmental regulators. *Genes Dev.* **7**, 2431–2445 (1993).
19. Kumari, D. & Usdin, K. Interaction of the transcription factors USF1, USF2, and α -Pal/Nrf-1 with the FMR1 promoter. Implications for Fragile X mental retardation syndrome. *J. Biol. Chem.* **276**, 4357–4364 (2001).
20. Spruijt, C. G. et al. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159 (2013).
21. Hu, S. et al. DNA methylation presents distinct binding sites for human transcription factors. *eLife* **2**, e00726 (2013).
22. Baubec, T., Ivanek, R., Lienert, F. & Schübeler, D. Methylation-dependent and -independent genomic targeting principles of the MBD protein Family. *Cell* **153**, 480–492 (2013).
23. Borgel, J. et al. Targets and dynamics of promoter DNA methylation during early mouse development. *Nature Genet.* **42**, 1093–1100 (2010).
24. Ficiz, G. et al. FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell* **13**, 351–359 (2013).
25. Habibi, E. et al. Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13**, 360–369 (2013).
26. Hon, G. C. et al. Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res.* **22**, 246–258 (2012).
27. ENCODE Project Consortium An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
28. Lienert, F. et al. Identification of genetic elements that autonomously determine DNA methylation states. *Nature Genet.* **43**, 1091–1097 (2011).
29. Krebs, A. R., Dessus-Babus, S., Burger, L. & Schübeler, D. High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions. *eLife* **3**, e04094 (2014).
30. Sherwood, R. I. et al. Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nature Biotechnol.* **32**, 171–178 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to S. Dessus-Babus, K. Jacobeit and T. Roloff (FMI) for processing deep-sequencing samples, to C. Wirbelauer for technical assistance and to A. Arnold for technical advice. We thank M. Stadler and D. Gaidatzis for bioinformatic advice and members of our laboratory, N. Thomae (FMI) and M. Lorincz (UBC Vancouver) for comments on the manuscript. We apologize to colleagues whose work we could not cite owing to space limitations. Research in the laboratory of D.S. is supported by the Novartis Research Foundation, the European Union (NoE ‘EpiGeneSys’ FP7-HEALTH-2010-257082 and the ‘Blueprint’ consortium FP7-282510), the European Research Council (EpiGePlas) and the Swiss initiative in Systems Biology (RTD Cell Plasticity). A.F.B. and P.A.G. are supported by EMBO postdoctoral long-term fellowships and S.D. and D.H. by predoctoral fellowships from the Boehringer Ingelheim Fonds.

Author Contributions A.F.B., L.B., S.D. and D.S. initiated and designed the study; S.D. performed the experiments; A.F.B. performed the data analysis; S.D. contributed to data analysis; S.D. and P.A.G. generated the TKO cell line; D.H. generated the overexpression construct; L.B. advised on data analysis; D.S. supervised all aspects of the project; the manuscript was prepared by S.D., A.F.B. and D.S. All authors discussed results and commented on the manuscript.

Author Information Genome-wide datasets generated for this study are deposited at GEO under the accession number GSE67867. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.S. (dirk@fmi.ch).

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The investigators were not blinded to allocation during experiments and outcome assessment.

Cell culture. Mouse ES cells HA36CB1/159-2 (denoted hereafter as 159) derived from mixed 129-C57Bl/6 background blastocysts²², TC-1 cells and REST knockout and corresponding wild-type cells^{31,32} were cultivated without feeders on 0.2% gelatine-coated dishes in DMEM, supplemented with 15% fetal calf serum, 1 × non-essential amino acids, 2 mM L-glutamine, LIF and 0.001% β-mercaptoethanol (37°C, 7% CO₂). Serum-free cultivation was performed in N2B27 medium, supplemented with 1 × non-essential amino acids, 2 mM L-glutamine, LIF and 0.001% β-mercaptoethanol, as well as MEK inhibitor PD0325901 (1 μM) and GSK3 inhibitor CHIR99021 (3 μM), together known as 2i. For switching between culturing conditions, cells were cultured for at least three weeks under the new conditions before performing downstream experiments. Mouse 159 ES cells were differentiated to neuronal progenitors as previously described³³. HMECs were purchased from Lonza (CC-2551), cultivated according to the supplier's instructions and collected after two passages. HCC1954 cells were cultured in RPMI 1640 medium supplemented with 10% fetal calf serum, 1 × nonessential amino acids and 1 × L-glutamine (37°C, 5% CO₂).

Generation of isogenic DNMT TKO cell lines. Mouse 159 ES cells were co-nucleofected with three plasmids expressing mammalian-codon optimized Cas9 and sgRNAs targeting the region coding for the active PCQ/N loop in *Dnmt1*, *Dnmt3a*, and *Dnmt3b* (parental vector pX330, guide oligo sequences: *Dnmt1* (CACCTGTGGTGGGCCACCCTGCCA, AAAGTGGCAGGGTGG CCCACCAACA), *Dnmt3a* (CACCGACAATGGAGAGGTCATTGC, AAACGCAATGACCTCTCCATTGTC), *Dnmt3b* (CACCGTTAGAGAG ATCATTCAT, AAACATGCAATGATCTCTCTAACG)). A plasmid conveying resistance against puromycin was co-transfected. Puromycin selection (2 μg ml⁻¹) was carried out one day after transfection for 48 h. After five days of recovery, individual colonies were picked and genotyped by methylation-sensitive HpaII digest, using methylation-insensitive MspI digest as control. For clones in which loss in methylation was observed, *Dnmt* genes were sequenced to confirm successful targeting of all six alleles. Global 5-methylcytosine and 5-hydroxy-methylcytosine levels in positive TKO clones were measured by Zymo Research (<http://www.zymoresearch.com>), using high-pressure liquid chromatography coupled to mass spectrometry.

RNA isolation. RNA was isolated with the RNeasy mini kit (Qiagen) with on-column DNA digestion. For RNA-seq, two micrograms of total RNA from three independent cultures were depleted from ribosomal RNA using the Ribo-Zero rRNA removal kit (Epicentre).

DNase footprinting. DNase treatment of wild-type and TKO cells was performed essentially as previously described, with some modifications³⁴. Briefly, intact nuclei were extracted using 0.03% NP-40 in an isotonic buffer. After NP-40 removal, batches of 5 million nuclei were incubated for 4 min at 37°C with a range of DNase I (DPRF, Worthington) concentrations in the presence of Ca²⁺. The digestion was stopped by addition of EDTA and SDS and the samples were treated with proteinase K and RNase A. Phenol-chloroform extracted DNA was separated on a 5–30% sucrose gradient by ultracentrifugation for 24 h and fractionated with a Gilson fraction collector FC 203B. Fractions were precipitated with ethanol and resuspended in TE buffer. Both successful digestion and size separation were verified by agarose gel electrophoresis. In addition, qPCR for amplicons within or outside known DHSs was used to confirm enrichment of DHSs in DNase-treated versus untreated and size-selected versus total DNA (primer sequences available upon request). Low-coverage sequencing of a barcoded pool of samples derived from different fractions of the sucrose gradient and treated with different DNase concentrations was used to select the sample with the highest information content. Based on this, the fraction of the gradient containing the shortest fragments (1–100 bp) was chosen for high-coverage sequencing.

Chromatin immunoprecipitation. Chromatin immunoprecipitation (ChIP) was carried out essentially as previously described³⁵, using a monoclonal antibody against NRF1 (Abcam, ab55744) and a polyclonal one against H3K27ac (Abcam, ab4729). ChIP–qPCRs were performed on at least three independent ChIP replicates according to standard protocols. Primer sequences are available upon request.

Knockdown by siRNA. TKO cells were reverse transfected with four preselected siRNAs targeting *Nrf1* (Qiagen, FlexiTube GeneSolution, GS18181) and Lipofectamine RNAiMax (Life Technologies) in three biological replicates, using the supplier's positive and negative controls (Qiagen, AllStars Mm Cell Death Control siRNA, SI04939025, AllStars Negative Control siRNA, SI03650318). To test knockdown efficiency, RNA was isolated after 72 h, reverse transcribed (PrimeScript, Takara) and *Nrf1* and *Gapdh* levels were determined according to standard protocols using predesigned TaqMan probes (Applied Biosystems, 4331182 and 4448489). Protein levels were measured by western blot on nuclear

extracts. The most efficient siRNA targeting *Nrf1* (Mm_Nrf1_7 FlexiTube siRNA, SI05183738) and the negative control siRNA were used for RNA-seq experiments. **Transient overexpression.** For transient overexpression, NRF1 was placed under the control of the CAG promoter. *Nrf1* cDNA was amplified from a random hexamer reverse transcription cDNA library (Superscript III, Invitrogen) generated from total RNA extracts and cloned into pL1-CAGGS-bio-MCS-polyA-11²². Primer sequences are available upon request. This plasmid was reverse transfected into mouse 159 ES cells using Lipofectamine 2000 (Invitrogen). ChIP was performed 12 h after transfection. Overexpression was verified by western blot on nuclear extracts.

Recombinase-mediated cassette exchange. DNA fragments to be inserted into the ectopic genomic site in TC-1 cells were amplified from genomic DNA and cloned into a plasmid containing a multiple cloning site flanked by two inverted L1 *Lox* sites. We inserted two endogenous NRF1 binding sites (chr8: 113,271,870–113,272,282 and chr8: 123,020,293–123,020,670 for Extended Data Fig. 8) as well as part of the *Mrap* promoter (chr16: 90,738,245–90,738,944 for Fig. 4a), into which we integrated an NRF1 motif with Quickchange PCR mutagenesis by replacing the T at position chr16: 90,738,825 with CATG. Primer sequences are available upon request. Both unmethylated plasmids and plasmids that were *in vitro* methylated with M.SssI (NEB) were used for the recombinase-mediated cassette exchange reaction³⁶. Complete *in vitro* methylation of the plasmids was confirmed by digestion with HpaII/MspI. Recombinase-mediated cassette exchange was performed in TC-1 ES cells as previously described^{28,35}. Single clones were picked 12 days after nucleofection and tested for successful insertion events by PCR. To remove methylation after insertion, clones were treated with 25 nM 5-Aza-2'-deoxycytidine (Sigma) for 4 days. For analysis of wild-type and mutated fragments of the *Gtf2a1* promoter, we used previously described clones that were generated in the same way²⁸.

Targeted amplicon bisulfite sequencing. For high coverage amplicon bisulfite sequencing of NRF1 binding sites target regions containing the highest confidence NRF1 motif (CGCATGCG) were selected based on high NRF1 ChIP enrichments in the TKO cell line, absence of enrichment in the wild-type and wild-type methylation levels of at least 80%. Primers for 200–400 bp amplicons were designed using our AmpliconBiSeq R package (<https://github.com/BIMSBbioinfo/AmpliconBiSeq>) and 56 pairs were randomly selected from this set. In addition, primers for 6 NRF1 motifs that were unbound in the TKO cell line, 9 unmethylated regions (UMRs), 9 fully methylated regions (FMRs), 9 constitutive REST/CTCF LMRs and T7/lambd were included as controls, resulting in 96 primer pairs in total (Supplementary Table 3). Primers were commercially synthesized in a 96-well plate format (Microsynth). Genomic DNA was isolated at the same time point as collection for ChIP. Bisulfite conversion was performed on 2 μg of the RNaseA-treated DNA mixed with 3.2 pM M.SssI methylated T7 and unmethylated lambda DNA as conversion controls (EpiTect Bisulfite kit, Qiagen). Bisulfite-converted DNA was amplified in a 96-well format with the designed specific primers using the following cycling conditions: 20 touch-down cycles from 55 to 50°C with 30 s at 95°C, 30 s at 55/50°C and 30 s at 72°C, followed by 36 cycles of 30 s at 95°C, 30 s at 50°C and 30 s at 72°C and a final 5 min extension step at 72°C. Then 5 μl of each individual PCR reaction were combined and the pool was size-selected using Agencourt AMPure XP beads (Beckman Coulter) before library preparation. Methylation profiling for insertions as well as REST motif-containing LMRs/UMR was performed with the same settings (genomic coordinates and primers in Supplementary Table 3).

Library preparation and next-generation sequencing. DNase-seq libraries were prepared essentially according to standard Illumina protocols, using 40 ng of the precipitated fractions of the sucrose gradient as starting material. To reduce amplification bias, end-repaired, A-tailed and adaptor ligated DNA was amplified in 6 cycles of PCR with KAPA HiFi Hot Start polymerase. Adaptor dimers were subsequently removed with Agencourt AMPure XP beads (Beckman Coulter). For sequencing of total RNA, strand-specific RNA-seq libraries were prepared from rRNA depleted samples using the ScriptSeq v2 protocol (Epicentre). Libraries for ChIP-seq were prepared according to standard Illumina library preparation protocols, with matching input sequenced for each IP. Twelve cycles of PCR (NEB Q5 Hot Start HiFi PCR) were performed on end-repaired, A-tailed and adaptor-ligated DNA before gel size-selection. Libraries for whole genome bisulfite sequencing were prepared essentially as previously described⁸. Briefly, 5 μg of sonicated genomic DNA were end repaired and 3'-end adenylated using the Illumina TruSeq DNA LT Sample Preparation kit (Illumina 15025064). Paired-end adapters were ligated to the DNA fragments and adaptor-ligated DNA was purified by 2% agarose gel electrophoresis. The gel-purified DNA was converted with the EpiTect bisulfite kit (Qiagen). Converted libraries were enriched by 10 cycles of PCR using PfuTurbo Cx Hotstart DNA Polymerase (Agilent) and purified using AMPure XP beads. For amplicon bisulfite sequencing, libraries of purified PCR pools were prepared according to standard Illumina library preparation protocols using 12 cycles of

PCR (NEB Q5 Hot Start HiFi PCR). Quality of the libraries and size distribution was assessed on an Agilent 2100 Bioanalyzer (Agilent Technologies). For RNA-seq, DNase-seq benchmarking, ChIP-seq and amplicon bisulfite sequencing, three to six samples with different barcodes were mixed at equimolar ratios per pool. Sequencing was performed on an Illumina HiSeq 2500 machine (DNase-seq, RNA-seq, ChIP-seq: 50 bp read length, single-end; whole-genome bisulfite sequencing: 100 bp read length, paired end) or a MiSeq machine (DNase-seq benchmarking: 25 bp read length, paired end; amplicon bisulfite sequencing: 250 bp read length, paired end) according to Illumina standards.

Sequencing data processing. RNA-seq reads were mapped to the mouse reference transcriptome (NCBIM37.67) using TopHat³⁷ version 1.3.1 with parameter `no-novel-juncs`. DNase-seq reads were trimmed for Illumina adaptors. DNase-seq and ChIP-seq reads were mapped to the mouse reference genome (mm9 only chromosomes 1 to 19, X, Y and M) or human reference genome (hg19 only chromosomes 1 to 22, X, Y and M) using Bowtie³⁸ version 1.0.0 with parameters `-v 3 -m 1 -best -strata`. Whole-genome Bis-seq reads were processed with QuasR³⁹ and positions covered by at least 10 reads were used. Amplicon bisulfite sequencing samples were analysed with the AmpliconBiSeq R package (<https://github.com/BIMSBbioinfo/AmpliconBiSeq>). Amplicons with at least 100× (TKO-specific NRF1 sites) or 30× (insertions) coverage were selected for downstream analysis.

Visualization of read densities. We used the first bp (5'-end) of the DNase-seq reads (DNase I cut site), the ChIP-seq reads extended to 200 bp (average estimated fragment length) and split RNA-seq reads to calculate the read density normalized to one million reads in the library for each genomic position (BigWig files). Screenshots of genomic regions were taken using the UCSC genome browser⁴⁰.

Identification of enriched regions. DHSs were identified as regions with enriched DNase I cuts using a sliding window approach. The mean read density for each region of 51 bp was calculated by steps of 10 bp within mappable regions and outside ENCODE blacklisted regions²⁷. Regions with a mean density of 0.001 (about 10 DNase I cuts) and at least 10 bp covered were merged and kept if their length was at least 100 bp. Enriched ChIP-seq regions over corresponding input were identified using the peak calling software Peakzilla⁴¹ with default parameters.

Correlation of read counts. We used the first bp (5'-end) of the DNase-seq reads (DNase I cut site), the ChIP-seq reads extended to 200 bp (average estimated fragment length) and split RNA-seq reads to calculate raw read counts for regions of interest (merged DNase-seq or ChIP-seq enriched regions or genes). The R package DESeq⁴² was used to normalize the raw read counts and identify differential regions using a fold change threshold of 2 and an adjusted *P* value threshold of 10^{-3} for DNase-seq and ChIP-seq regions and 10^{-5} for RNA-seq data sets. We generated scatterplots and calculated Pearson correlation coefficients (PCC) from the normalized read counts using R.

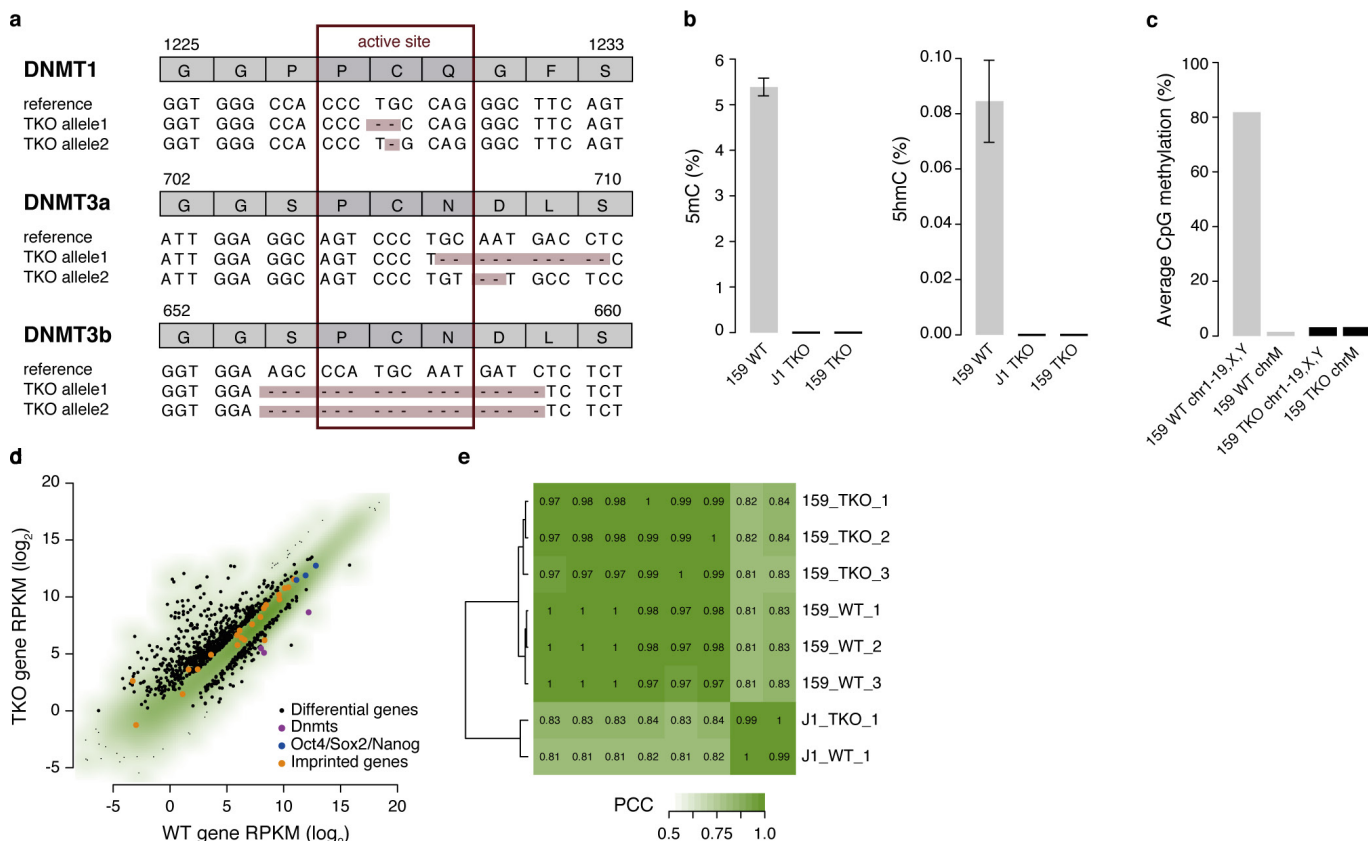
Functional analyses. Germline-specific imprinted regions were used from ref. 43. Peaks were assigned to their closest gene transcriptional start site (TSS) using the mouse reference transcriptome (NCBIM37.67) and human reference transcriptome (GRCh37.71). The conservation rate of regions was calculated using the PhastCons 11 way placental mammals⁴⁴.

Motif-enrichment analysis. We searched DHS regions for known motifs from JASPAR⁴⁵, ref. 46 and UniPROBE⁴⁷ using MAST⁴⁸ (from the MEME suite programs version 4.1.1) with a *P* value threshold of 2.44×10^{-4} ($(0.25)^6$) (see Supplementary Table 1). The statistical significance of the differential motif enrichment was assessed by a hypergeometric *P* value.

Published data sets. RNA-seq data sets in J1 mouse ES cells were obtained from GEO with the accession numbers GSM727427 and GSM727428 (ref. 16), in mouse ES cells cultured in serum from GSM590126, GSM758167 and GSM758168 (ref. 49), and 2i from GSM758168, GSM590128 and GSM590129 (ref. 49), in neuronal progenitors from GSM778489 and GSM778490 (ref. 50), in HMEC cells from GSM721141 (ref. 26), in HCC1954 cells from GSM721140 (ref. 26), in h1hESC from GSM758566 (ref. 51) and in GM12878 from GSM758559 (ref. 51). DNase-seq data sets in mouse ES cells were obtained from GSM1014159 (ref. 51). Bis-seq

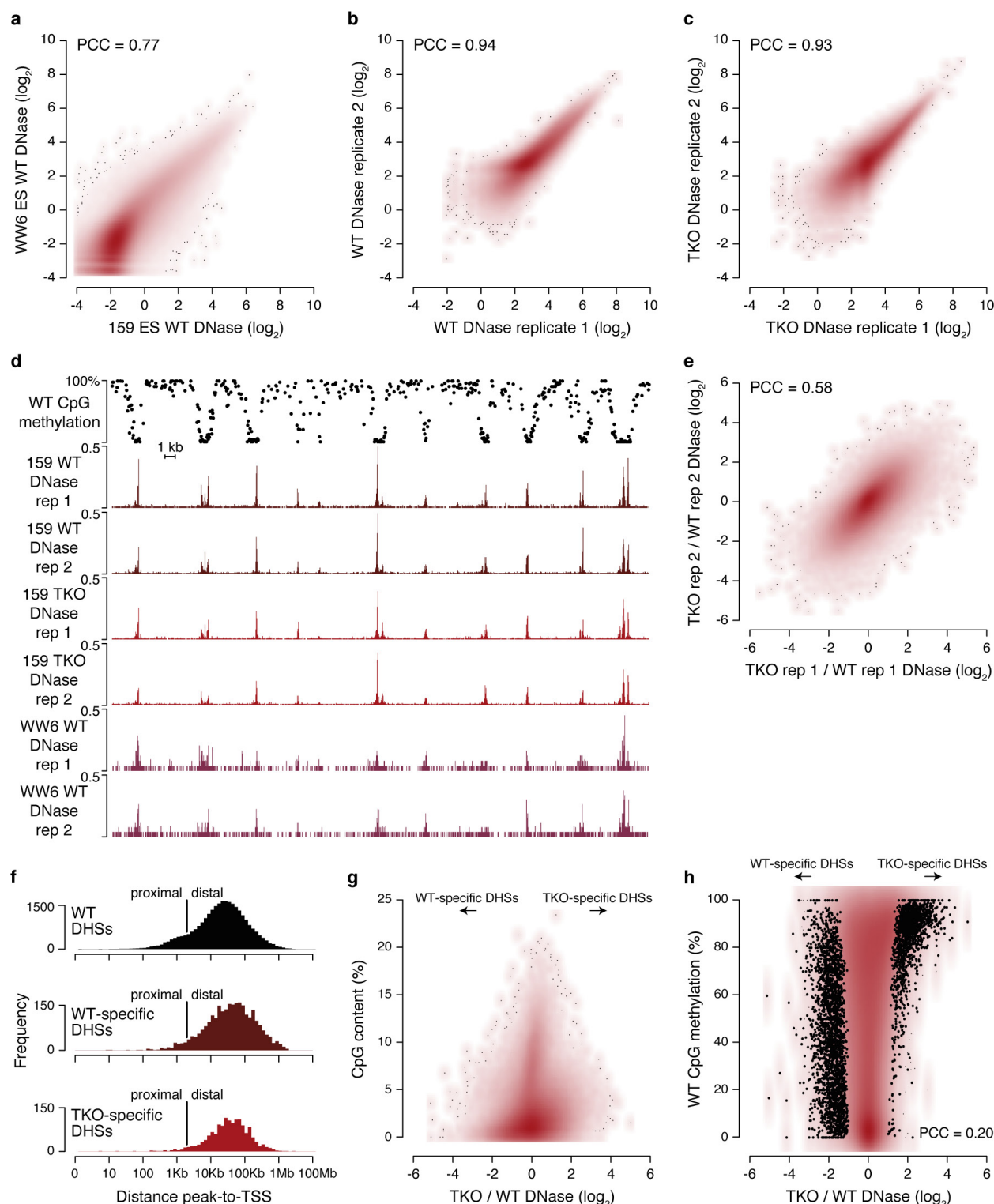
data sets in mouse ES cells were obtained from GSM748786 (ref. 8), in neuronal progenitors from GSM748788 (ref. 8), in HMEC cells from GSM721195 (ref. 26), in HCC1954 cells from GSM721194 (ref. 26), in H1-hESC from GSM1002649 (ref. 27) and in GM12878 from GSM1002650 (ref. 27). ChIP-seq data sets were obtained for NRF1 in H1-hESC from GSM935308 (ref. 27) and in GM12878 from GSM935309 (ref. 27), in mouse ES cells for MeCP2 from GSM972976 (ref. 22), for CTCF from GSM747534 (ref. 8), for REST from GSM671094 (ref. 52), for ZFX from GSM288352 (ref. 53), for KLF4 from GSM288354 (ref. 53), for ESRRB from GSM288355 (ref. 53), for cMYC from GSM288356 (ref. 53), for nMYC from GSM288357 (ref. 53), for OCT4 from GSM307137 (ref. 54), for SOX2 from GSM307138 (ref. 54) and for NANOG from GSM307141 (ref. 54).

31. Jørgensen, H. F., Chen, Z. -F., Merckenschlager, M. & Fisher, A. G. Is REST required for ESC pluripotency? *Nature* **457**, E4–E5, E7 (2009).
32. Chen, Z. F., Paquette, A. J. & Anderson, D. J. NRSF/REST is required *in vivo* for repression of multiple neuronal target genes during embryogenesis. *Nature Genet.* **20**, 136–142 (1998).
33. Bibel, M., Richter, J., Lacroix, E. & Barde, Y.-A. Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nature Protocols* **2**, 1034–1043 (2007).
34. John, S. et al. Genome-scale mapping of DNase I hypersensitivity. *Curr. Protoc. Mol. Biol.* **Chapter 27**, Unit 21.27–21.27.20 (2013).
35. Jermann, P., Hoerner, L., Burger, L. & Schübeler, D. Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation. *Proc. Natl Acad. Sci. USA* **111**, E3415–E3421 (2014).
36. Schübeler, D. et al. Genomic targeting of methylated DNA: influence of methylation on transcription, replication, chromatin structure, and histone acetylation. *Mol. Cell. Biol.* **20**, 9103–9112 (2000).
37. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-seq. *Bioinformatics* **25**, 1105–1111 (2009).
38. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
39. Gaidatzis, D., Lerch, A., Hahne, F. & Stadler, M. B. QuasR: quantification and annotation of short reads in R. *Bioinformatics* **31**, 1130–1132 (2015).
40. Kent, W. J. et al. The Human Genome Browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
41. Bardet, A. F. et al. Identification of transcription factor binding sites from ChIP-seq data at high resolution. *Bioinformatics* **29**, 2705–2713 (2013).
42. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
43. Xie, W. et al. Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* **148**, 816–831 (2012).
44. Siepel, A. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
45. Sandelin, A. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.* **32**, D91–D94 (2004).
46. Jolma, A. et al. DNA-binding specificities of human transcription factors. *Cell* **152**, 327–339 (2013).
47. Newburger, D. E. & Bulyk, M. L. UniPROBE: an online database of protein binding microarray data on protein–DNA interactions. *Nucleic Acids Res.* **37**, D77–D82 (2009).
48. Bailey, T. L. & Gribskov, M. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* **14**, 48–54 (1998).
49. Marks, H. et al. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590–604 (2012).
50. Tippmann, S. C. et al. Chromatin measurements reveal contributions of synthesis and decay to steady-state mRNA levels. *Mol. Syst. Biol.* **8**, 593 (2012).
51. Yue, F. et al. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355–364 (2014).
52. Arnold, P. et al. Modeling of epigenome dynamics identifies transcription factors that mediate Polycomb targeting. *Genome Res.* **23**, 60–73 (2013).
53. Chen, X. et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).
54. Marson, A. et al. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**, 521–533 (2008).



Extended Data Figure 1 | Characterization of an isogenic DNMT TKO cell line created with CRISPR/Cas9. **a**, Frameshift deletions (brown) introduced at the active PCQ/N loops of the three DNA methyltransferases by CRISPR/Cas9 genome editing. **b**, Levels of 5-methyl-C and 5-hydroxy-methyl-C in the wild-type, isogenic (mouse ES cell line 159) and traditional (J1) TKO cell lines as determined by mass spectrometry. **c**, Average CpG methylation in wild-type and TKO cell lines determined by whole-genome bisulfite sequencing. Methylation in the TKO cell line is comparable to background levels represented by the methylation in chromosome M. **d**, Gene expression levels (RPKM) in isogenic wild type and TKO (159). Black dots represent significantly differentially expressed

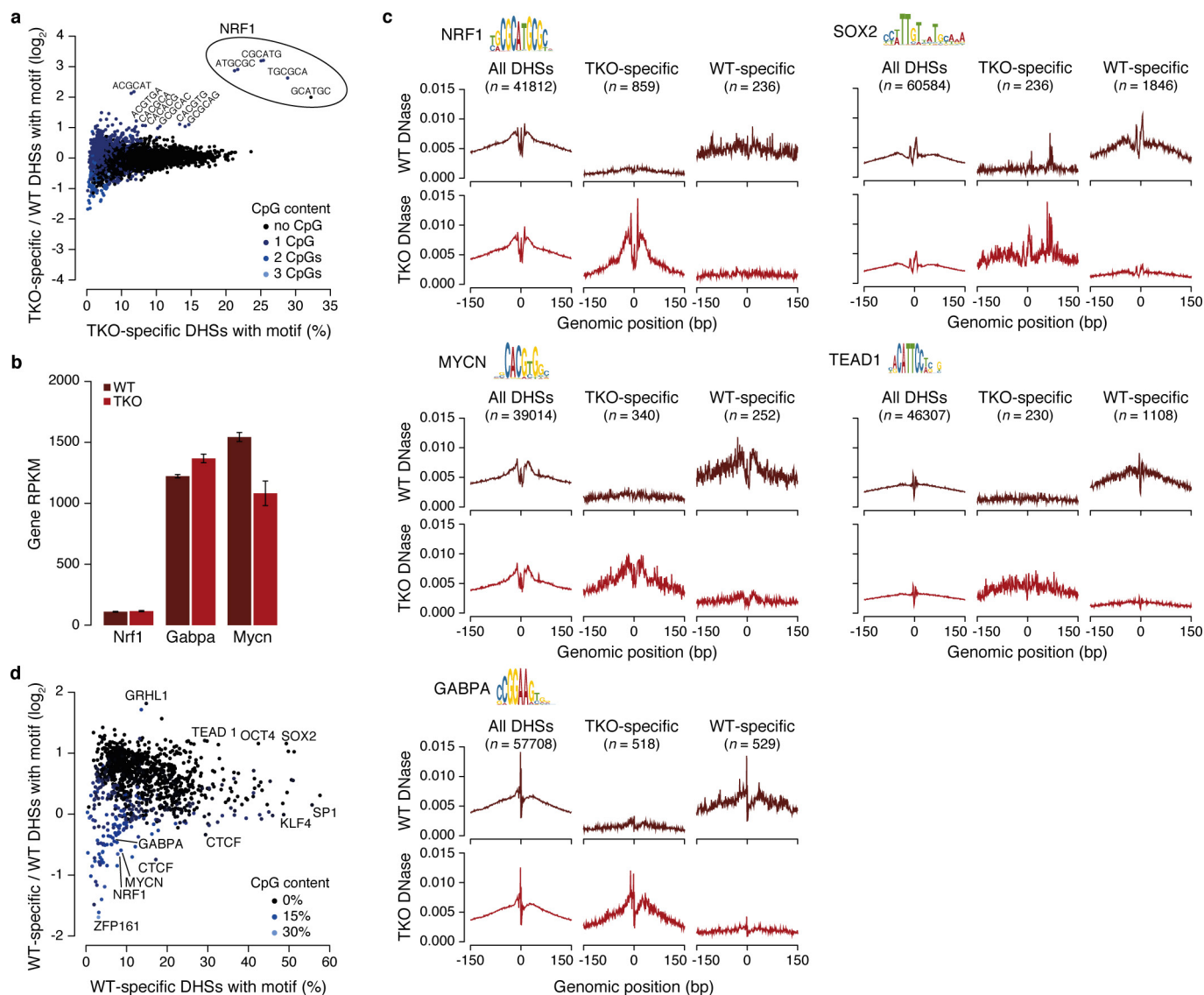
genes in wild type or TKO, with expected unregulation of germline genes¹⁶. The *Dnmt* genes are among the most downregulated genes (purple), while the majority of genes that reside within imprinted domains are upregulated roughly twofold (orange). Prominent marker genes of ES cells (*Oct4*, *Sox2* and *Nanog*, blue) remain unaltered. **e**, Hierarchical clustering of gene expression correlations for three independent 159 ES cell line wild-type and TKO replicates, and published J1 wild-type and TKO RNA-seq samples¹⁶. Overall, gene expression clusters by strain rather than presence of DNA methylation. This reflects the strong influence of genetic background on the global gene expression program and supports our approach of focusing further analysis on the isogenic TKO.



Extended Data Figure 2 | Characteristics of DNase-hypersensitive sites.

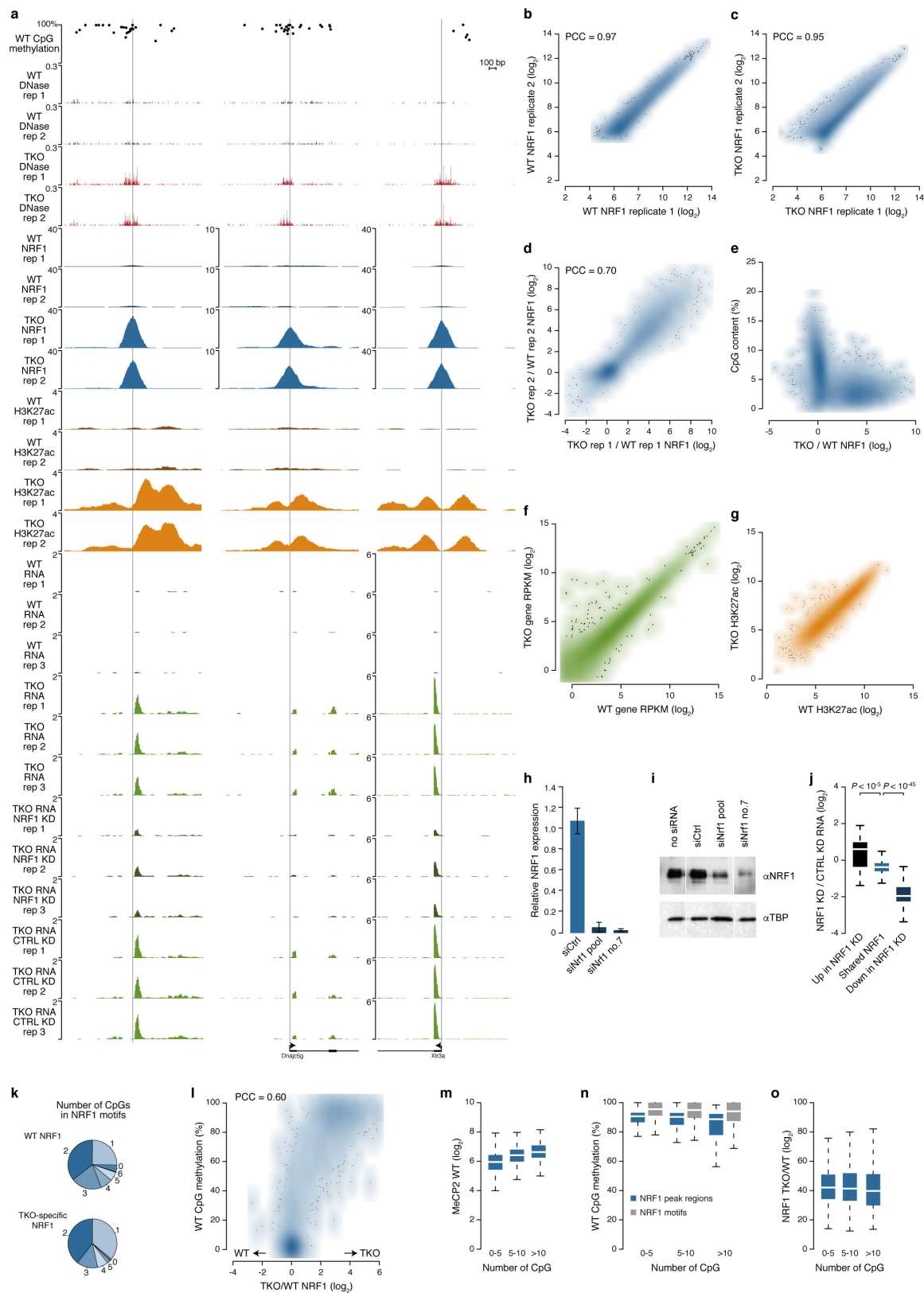
a, DNase-seq signal in our 159 ES cell line (wild-type) and an ENCODE WW6 ES cell (wild-type) DNase-seq sample²⁷ using a tiling window (500 bp) over the whole genome in mappable regions not blacklisted by ENCODE, illustrating that our protocol for genome-wide detection of DHSs matches available data sets in mouse ES cells. PCC was calculated on all DHSs. **b**, **c**, DNase-seq signal and PCC at all DHSs for independent biological replicates of wild type (**b**) and TKO (**c**). **d**, Wild-type methylation and replicates for DNase-seq signal in the 159 ES cell line (wild-type and TKO) and ENCODE WW6 (wild-type) at the genomic region from Fig. 1a (chr17: 25,920,000–25,972,499), illustrating that most DHSs remain unchanged upon removal of DNA methylation, in agreement with the overall similarity in gene expression. **e**, Change in

DNase-seq signal and PCC between wild type and TKO using different replicate samples, illustrating a high reproducibility of quantitative DHS changes between wild type and TKO. **f**, Distance of all wild-type, wild-type-specific or TKO-specific DHSs from closest gene transcriptional start site (TSS). Proximal and distal separation is at 2 kb. **g**, Change in DNase-seq signal between TKO and wild-type as a function of CpG content for all wild-type and TKO DHSs, illustrating that most changes occur in CpG-poor regions. **h**, Change in DNase-seq signal between TKO and wild-type versus average CpG methylation of all wild-type and TKO DHSs matching Fig. 1c, showing that TKO-specific DHSs (right) lie in regions with high methylation in wild type. Black dots represent significantly enriched DHSs (see Methods) in wild type ($n = 2,837$) or TKO ($n = 1,543$) from Fig. 1b.



Extended Data Figure 3 | Motif enrichment in cell-line-specific DNase-hypersensitive sites. **a**, Occurrence of all possible hexamers in TKO-specific DHSs compared to all wild-type DHSs. Blue colouring illustrates hexamer CpG content. Hexamers representing the NRF1 motif are highlighted by a circle. Most strongly enriched hexamers are labelled (only one of two reverse complements). **b**, Gene expression levels (RPKM) of candidate methylation-sensitive TFs in wild type and TKO indicating that differential abundance does not account for DHS formation upon loss of DNA methylation. Error bars are standard deviation from three biological replicates. **c**, Footprints of candidate TF motifs enriched in TKO-specific

(NRF1, MYCN, GABPA) or wild-type-specific (SOX2, TEAD1) DHSs shown as metaplot of wild-type (brown) or TKO (red) DNase-seq signal for all motifs in all wild-type and TKO (left), TKO-specific (middle) and wild-type-specific (right) DHSs. Number of regions is indicated above each metaplot. A DNase footprint is apparent at the NRF1 motif and, to a lesser extent, at MYCN and GABPA motifs specifically in TKO-specific sites in the TKO sample, whereas footprints at SOX2 and TEAD1 motifs in wild-type-specific sites are less unique to that cell state. **d**, Motif occurrences in wild-type-specific DHSs compared to all wild-type DHSs. Blue colouring illustrates motif CpG content.



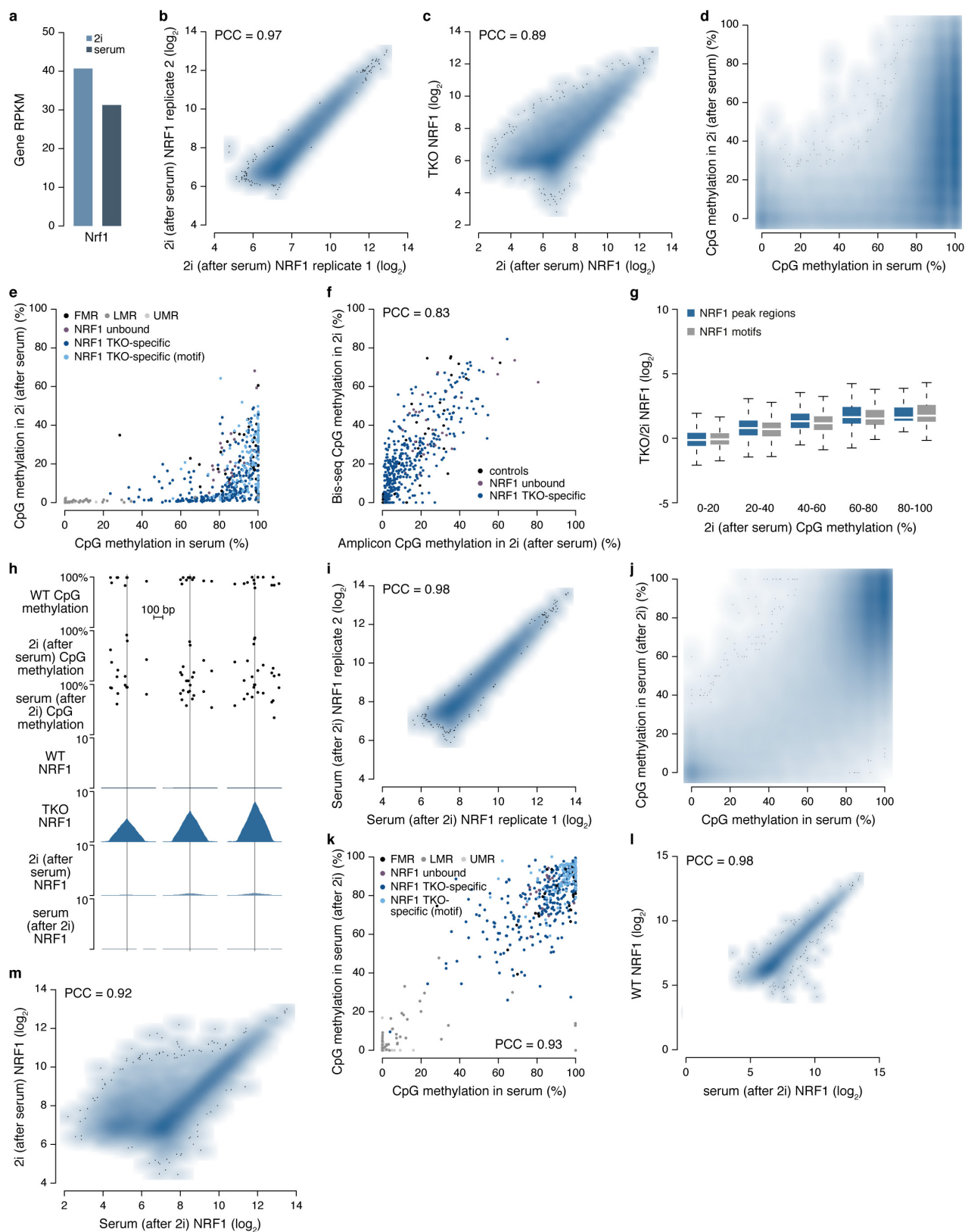
Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | Characteristics of NRF1 binding sites.

a, Wild-type methylation, and wild-type and TKO DNase-seq, NRF1 ChIP-seq, H3K27ac ChIP-seq and RNA-seq signal also upon *Nrf1* and mock knockdown in TKO at TKO-specific distal (left, chr4: 99,235,170–99,237,170; from Fig. 2a) and proximal (middle, chr5: 31,409,700–31,411,700; right, chrX: 70,341,500–70,343,500) genomic regions. The transcripts initiated directly at the NRF1 binding sites in TKO cells are specifically reduced upon knockdown of *Nrf1*, implying that they are indeed NRF1-dependent. **b, c**, NRF1 ChIP-seq signal at all NRF1 peak regions for independent biological replicates of wild type (**b**) and TKO (**c**). **d**, Change in NRF1 ChIP-seq signal and PCC between wild type and TKO using different replicate samples, illustrating a high reproducibility of quantitative NRF1 changes between wild type and TKO. **e**, Change in NRF1 ChIP-seq signal between TKO and wild type versus CpG content of all wild-type and TKO NRF1 peak regions, illustrating that most changes occur in CpG-poor regions. **f**, RNA expression levels (RPKM) in wild type and TKO at all wild-type and TKO NRF1 peak regions, illustrating the appearance of a few aberrant TKO-specific transcripts directly at NRF1 binding sites. **g**, H3K27ac ChIP-seq signal in wild type and TKO at all wild-type and TKO NRF1 peak regions, illustrating appearance of TKO-specific acetylation at a few NRF1 binding sites. **h**, Knockdown efficiency for the pool of three siRNAs and most efficient single siRNA targeting *Nrf1* in TKO cells. Mean of three independent biological replicates normalized to GAPDH; error bars reflect standard deviation. Genetic deletion of *Nrf1* with CRISPR/Cas9 was lethal (data not shown). **i**, Reduction in nuclear NRF1 levels upon siRNA knockdown with pool of three siRNAs and most efficient single siRNA targeting *Nrf1* as measured

by western blot. Blot was cropped for clarity, all samples were loaded on the same gel (for uncropped gels see Supplementary Fig. 1). **j**, Expression change (in RPKM) of genes closest to shared and TKO-specific NRF1 peaks between TKO cells treated either with negative control siRNA or the most efficient single siRNA targeting *Nrf1*, showing highly significant loss in expression after knockdown. *P* values from Wilcoxon tests.

k, Number of CpGs in NRF1 motifs closest to peak summit in all wild-type (top) or TKO-specific (bottom) NRF1 peaks, illustrating that motifs in TKO-specific NRF1 peaks contain at least one CpG. **l**, Change in NRF1 ChIP-seq signal between TKO and wild type versus average methylation in wild type at all NRF1 sites corresponding to Fig. 2g, illustrating that increased NRF1 binding in TKO occurs at regions that were methylated in wild type. **m–o**, Average wild-type MeCP2 ChIP-seq signal²² (**m**), wild-type methylation in NRF1 peak regions or in NRF1 motifs closest to peak summits (**n**) and change of NRF1 signal between wild type and TKO (**o**) within 500 bp regions around TKO-specific NRF1 peak summits grouped according to CpG density (0–5 CpGs, *n* = 3,680; 5–10 CpGs, *n* = 2,477; >10 CpGs, *n* = 680). If indirect repression could contribute to differential NRF1 binding, we would expect a more pronounced increase of NRF1 binding at sites with higher CpG density upon demethylation of the genome, as methyl-CpG binding domain proteins (MBDs) such as MeCP2 bind preferentially to regions with a high density of methylated CpGs rather than fully methylated regions with low CpG density. TKO-specific binding of NRF1 is independent of CpG density and MeCP2 enrichment in the methylated genome, strongly arguing against an involvement of indirect repression in NRF1 binding site restriction.

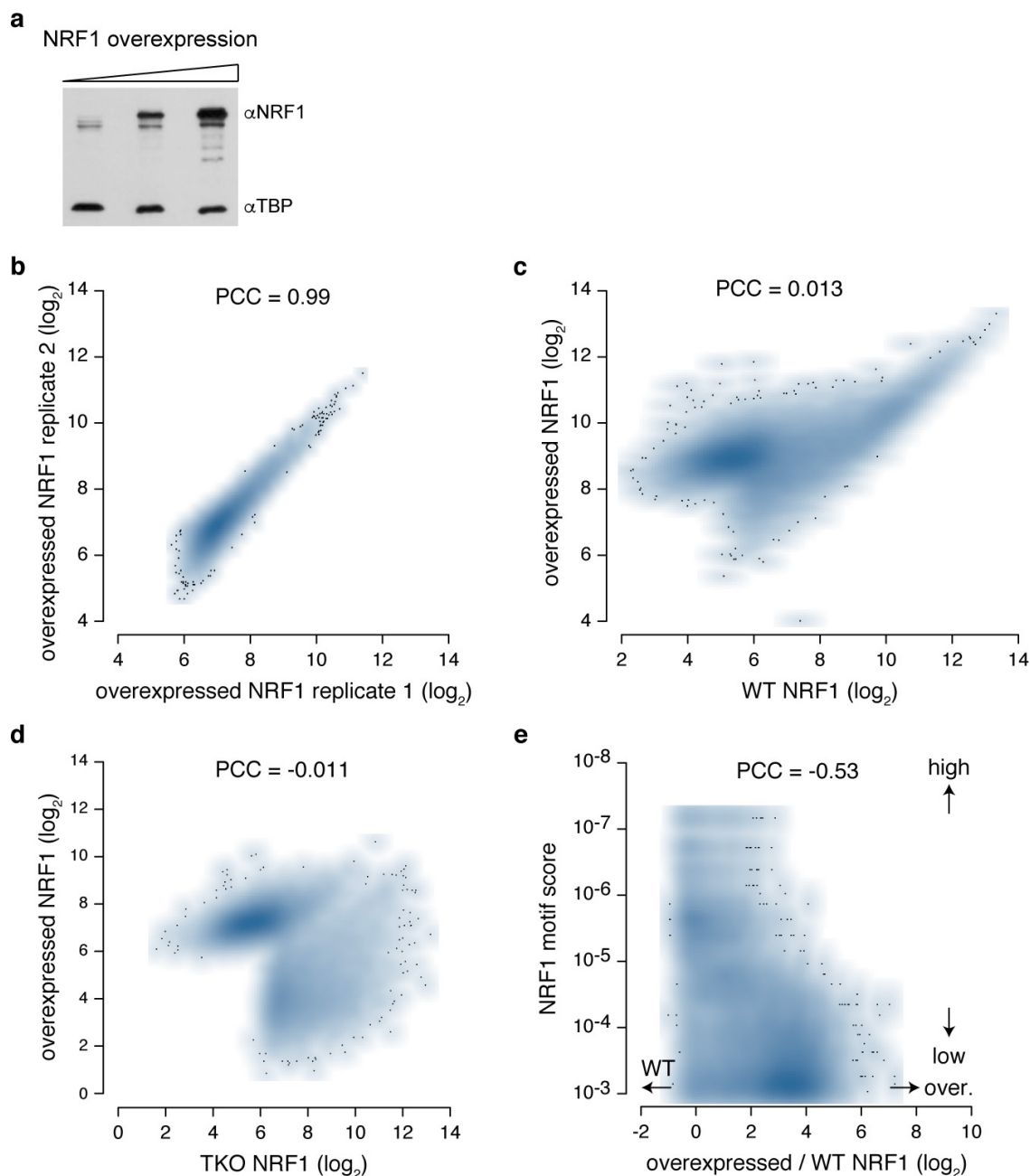


Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | NRF1 binding in different culture conditions.

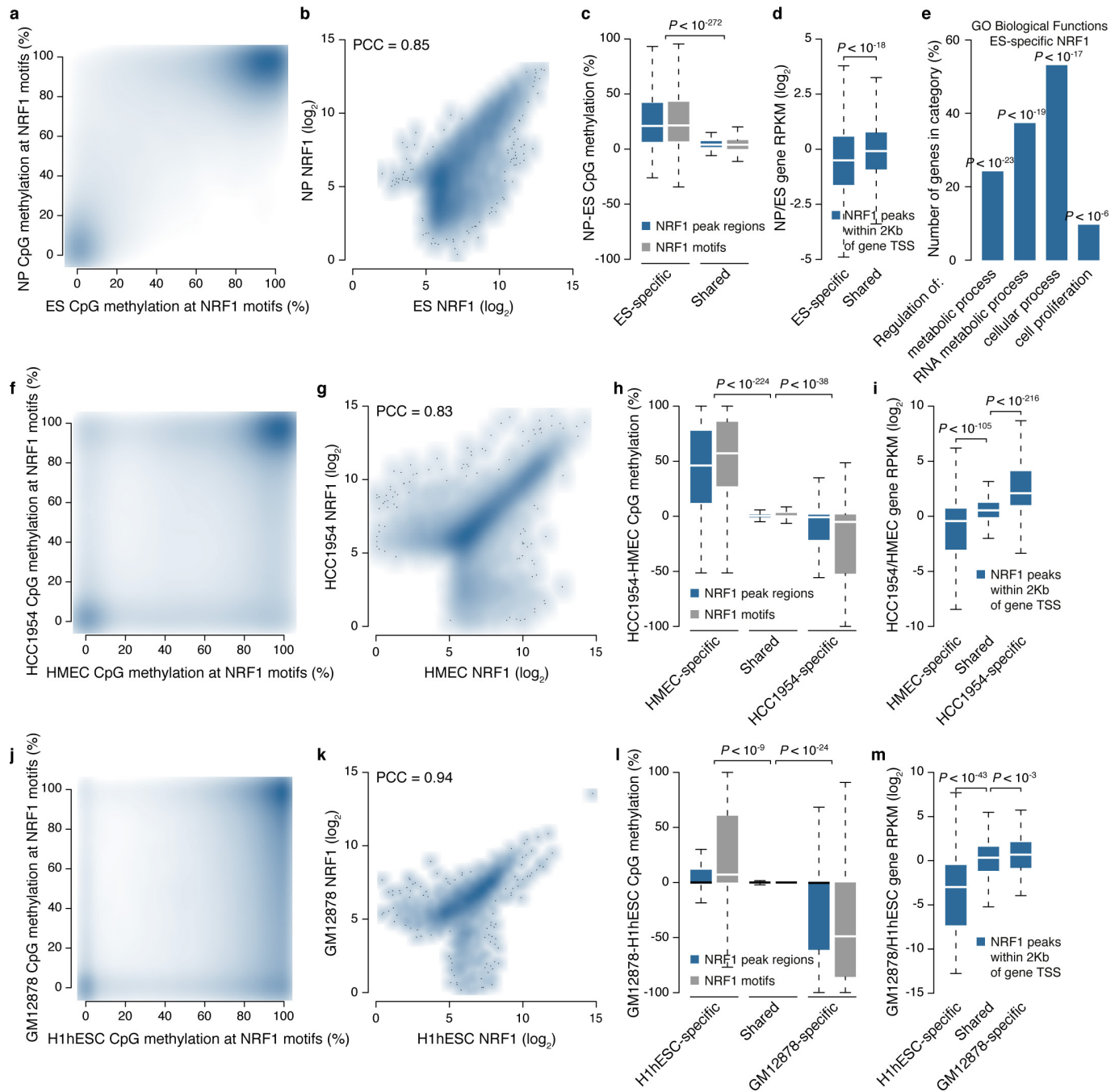
a, *Nrf1* gene expression levels (RPKM) in 2i and serum culture conditions⁴⁹. **b**, NRF1 ChIP-seq signal in wild-type cells adapted to 2i culture conditions (after culture with serum) for two biological replicates. **c**, NRF1 ChIP-seq signal in wild-type cells adapted to 2i (after culture with serum) and TKO. **d**, Methylation in wild-type cells cultured in serum and 2i (after culture with serum) at all NRF1 motifs. **e**, Methylation in serum and 2i (after culture with serum) measured by amplicon Bis-seq for fully methylated (FMR), low methylated (LMR), unmethylated (UMR) controls, 6 unbound NRF1 sites and 56 TKO-specific NRF1 sites. **f**, Comparison and PCC of DNA methylation levels by amplicon Bis-seq and whole-genome Bis-seq upon culture in 2i (after culture with serum). **g**, Average 2i (after culture with serum) methylation in NRF1 peak regions or NRF1 motifs within peaks versus change in NRF1 signal between TKO and 2i (after culture with serum) at all NRF1 peaks, illustrating that reduced NRF1 binding in 2i compared to TKO can be explained by residual methylation. **h**, Methylation in wild-type cells cultured in serum, cultured in 2i (after culture with serum) and cultured in serum (after culture in 2i)

and NRF1 ChIP-seq signal in wild type, TKO, cultured in 2i (after culture with serum) and cultured in serum (after culture with 2i) at TKO-specific regions with higher 2i methylation in NRF1 motifs (grey lines) than surrounding region (left, chr10: 66,251,100–66,251,700; middle, chr4: 15,976,050–15,976,650; right, chr19: 55,833,420–55,834,020). NRF1 is unable to bind if CpGs in the motif remain methylated in 2i, even if the surrounding region is unmethylated. **i**, NRF1 ChIP-seq signal in wild-type cells adapted back to serum (after culture with 2i) for two biological replicates. **j**, Methylation in wild-type cells cultured in serum and adapted back to serum (after culture with 2i) at all NRF1 motifs. **k**, Methylation in wild-type cells cultured in serum and adapted back to serum (after culture with 2i) measured by amplicon Bis-seq for FMR, LMR and UMR controls, 6 unbound NRF1 sites and 56 TKO-specific NRF1 sites. **l**, NRF1 ChIP-seq signal in wild-type cells adapted back to serum (after culture with 2i) and original serum conditions. **m**, NRF1 ChIP-seq signal in wild-type cells adapted back to serum (after culture with 2i) and adapted to 2i (after culture with serum).



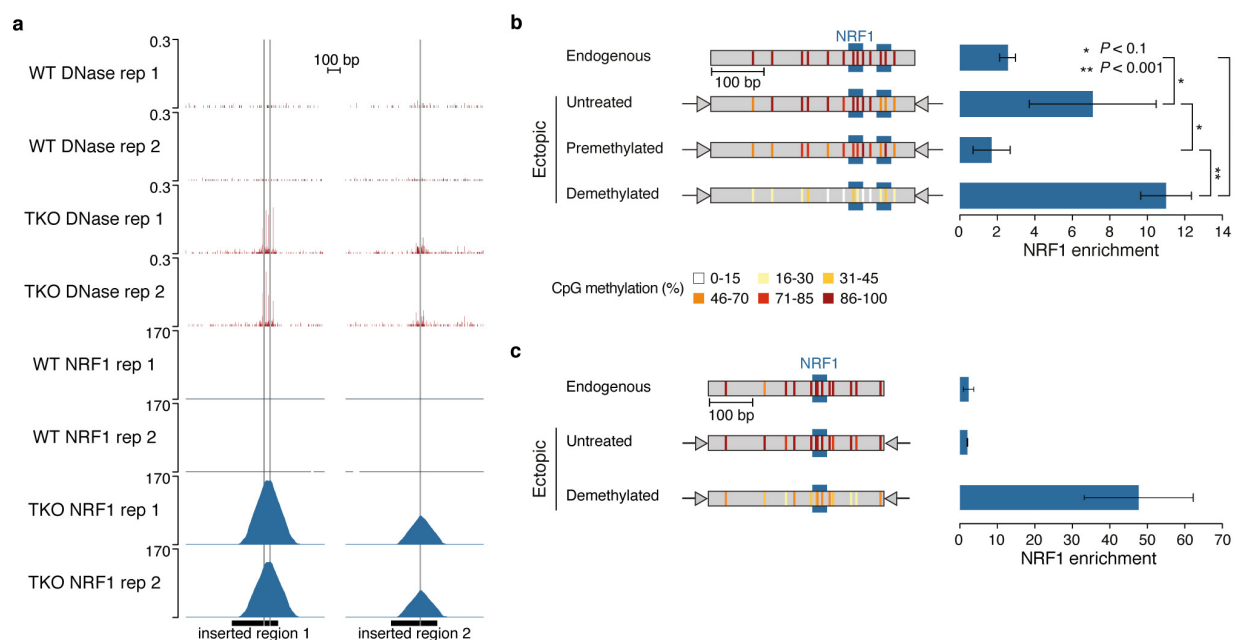
Extended Data Figure 6 | Overexpression of NRF1 is unable to induce binding to TKO-specific sites. **a**, Transient overexpression of NRF1 under control of the CMV (middle) or CAG promoter (right, used for ChIP experiments) leads to strong increase in nuclear NRF1 protein levels compared to endogenous levels (left) as measured by western blot (for uncropped gel data see Supplementary Fig. 1). The overexpressed protein contains a protein tag accounting for the higher molecular weight. **b**, NRF1 ChIP-seq signal upon transient NRF1 overexpression for two

biological replicates. **c**, NRF1 ChIP-seq signal in wild type and upon overexpression. **d**, NRF1 ChIP-seq signal in TKO and overexpression conditions only at TKO- and overexpression-specific NRF1 peak regions, illustrating that TKO-specific NRF1 sites are distinct from overexpression-specific sites. **e**, Change in NRF1 ChIP-seq signal between overexpression and wild type versus the score (MAST position *P* value) of NRF1 motifs closest to the summit, illustrating that sites gaining most NRF1 upon overexpression do not contain high-confidence motifs.



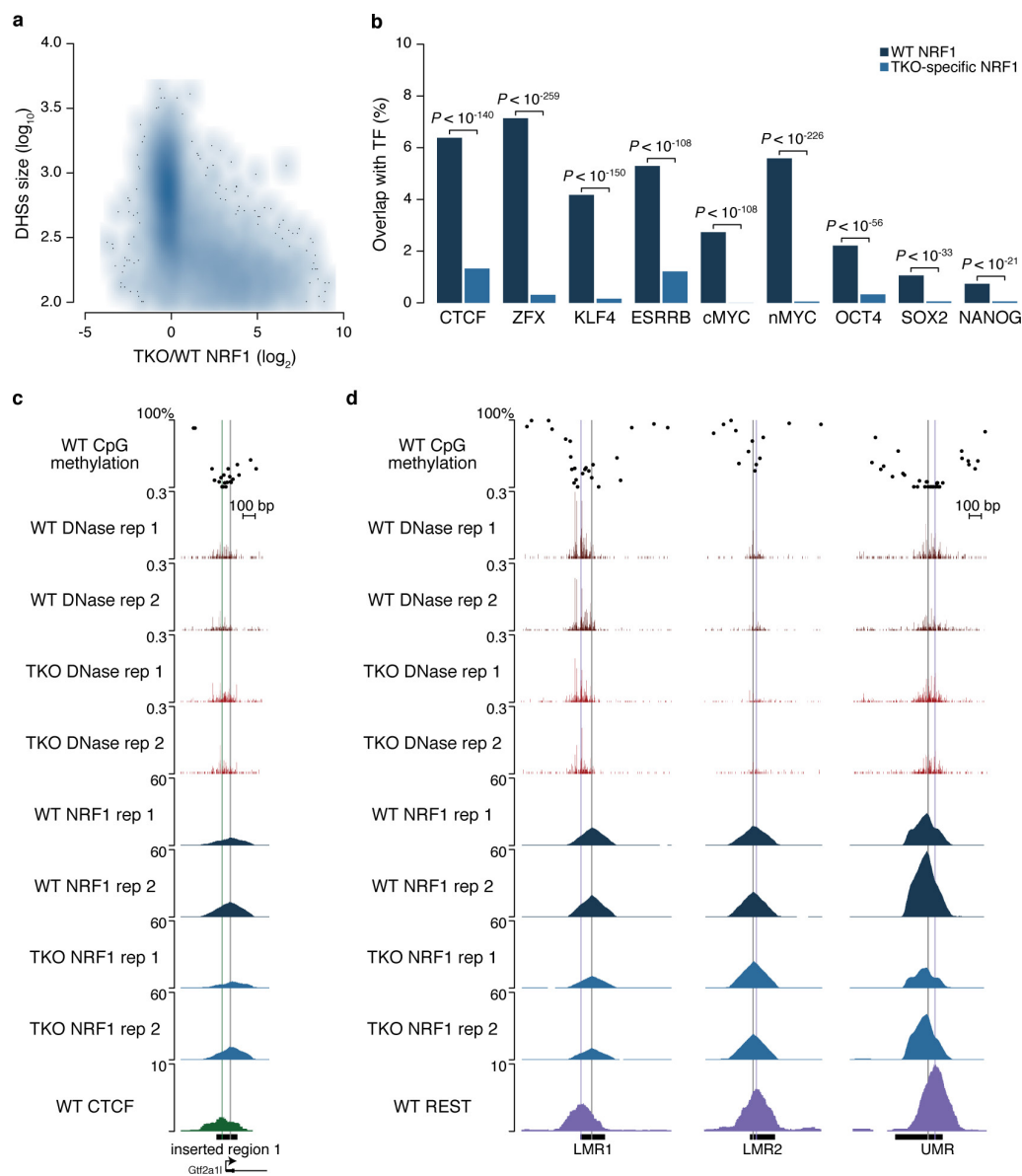
Extended Data Figure 7 | Cell-type-specific binding of NRF1 correlates with methylation and expression changes. **a–e**, Comparison of NRF1 binding in ES and neuronal progenitor cells. Methylation in ES and neuronal progenitors⁸ at all NRF1 motifs (**a**), NRF1 ChIP-seq signal in ES and neuronal progenitors at all NRF1 peaks (**b**), neuronal progenitor minus ES methylation of peak regions or NRF1 motifs in ES-specific ($n = 4,934$) and shared ($n = 4,951$) NRF1 peaks (negligible number of neuronal-progenitor-specific peaks) (**c**), expression of the genes⁵⁰ closest to ES-specific and shared NRF1 peaks (**d**), selection of gene ontology (GO) biological functions enriched in genes closest to ES-specific and shared NRF1 peaks (**e**). P values from Wilcoxon tests. **f–i**, Comparison of NRF1 binding in HMEC and HCC1954 cells. Methylation in HMEC and

HCC1954²⁶ at all NRF1 motifs (**f**), NRF1 ChIP-seq signal in HMEC and HCC1954 at all NRF1 peaks (**g**), HCC1954 minus HMEC methylation of peak regions or NRF1 motifs in HMEC-specific ($n = 2,726$), HCC1954-specific ($n = 2,685$) and shared ($n = 12,180$) NRF1 peaks (**h**), expression of the genes²⁶ closest to HMEC-specific, HCC1954-specific and shared NRF1 peaks (**i**). **j–m**, Comparison of NRF1 binding in H1-hESC and GM12878 cells. Methylation in H1-hESC and GM12878²⁷ at all NRF1 motifs (**j**), NRF1 ChIP-seq signal in H1-hESC and GM12878²⁷ at all NRF1 peaks (**k**), GM12878 minus H1-hESC methylation of peak regions or NRF1 motifs in H1-hESC- ($n = 618$), GM12878-specific ($n = 561$) and shared ($n = 3,198$) NRF1 peaks (**l**), expression of the genes²⁷ closest to H1-hESC-specific, GM12878-specific and shared NRF1 peaks (**m**).



Extended Data Figure 8 | NRF1 binding to the unmethylated motif can be recapitulated at an ectopic site. **a**, Wild-type and TKO DNase-seq and NRF1 ChIP-seq signal for two biological replicates at the endogenous counterparts of the inserted regions profiled in Extended Data Fig. 8b (left, chr8: 123,019,920–123,021,030) and Extended Data Fig. 8c (right, chr8: 113,271,460–113,272,690). **b**, Methylation (amplicon Bis-seq, left, coloured lines indicate position and methylation status of CpGs) and NRF1 binding (ChIP-qPCR, right) for an endogenous methylation-dependent NRF1 site (chr8: 123,020,293–123,020,670) and upon insertion of this region into a defined ectopic genomic locus. The position of the two NRF1 motifs containing two CpGs each is indicated in blue. The reporter construct was inserted either unmethylated or *in vitro* premethylated with M.SssI. In the untreated construct one motif becomes completely methylated upon insertion, whereas the other only gains roughly 50% methylation, and NRF1 binding is detected. The pre-methylated construct maintains at least one CpG with almost complete methylation in both core motifs present and shows strongly reduced NRF1 binding by comparison. Thus, the methylation sensitivity of NRF1 can

be recapitulated in an ectopic site even in the absence of global changes in DNA methylation. As expected, forcing complete demethylation of both core motifs in the premethylated insert by treatment of the cells with 5-aza-2'-deoxycytidine leads to further increased NRF1 binding compared to the untreated template. ChIP-qPCR enrichments are the mean of three independent biological replicates; error bars reflect standard deviation. See Supplementary Table 3 for methylation source data. **c**, Methylation (amplicon Bis-seq, left, coloured lines indicate position and methylation status of CpGs) and NRF1 binding (ChIP-qPCR, right) for an endogenous methylation-dependent NRF1 site (chr8: 113,271,870–113,272,282) and upon insertion of this region into a defined ectopic genomic locus. The untreated template gains full methylation in the core motif (blue) and does not show detectable NRF1 binding. Forcing complete demethylation by treatment with 5-aza-2'-deoxycytidine enables NRF1 to bind the site in the ectopic locus. ChIP-qPCR enrichments are mean of three independent biological replicates; error bars reflect standard deviation. See Supplementary Table 3 for methylation source data.



Extended Data Figure 9 | Constitutive NRF1 sites are co-bound by other TFs. **a**, Change in NRF1 ChIP-seq signal between TKO and wild type versus size of DHSs overlapping NRF1 peak regions, illustrating that wild-type NRF1 sites tend to overlap with larger DHSs. **b**, Overlap of wild-type and TKO-specific NRF1 peak regions with published ChIP-seq peak regions from other TFs expressed in ES cells^{8,53,54}, illustrating that wild-type NRF1 sites coincide with other TF binding events. *P* values from hypergeometric tests. **c**, Wild-type methylation, wild-type and TKO DNase-seq, and NRF1 and CTCF⁸ ChIP-seq signal for

two biological replicates at the endogenous *Gtf2a11* promoter (chr17: 89,067,600–89,068,350). The region used for the insertion experiments in Fig. 4b is indicated below. **d**, Wild-type methylation, wild-type and TKO DNase-seq for two biological replicates and NRF1 and REST⁵² ChIP-seq signal at adjacent NRF1 and REST binding sites (left, chr15: 100,703,260–100,704,500; middle, chr2: 180,152,200–180,153,150; right, chr2: 118,604,800–118,605,900). Regions profiled with amplicon Bis-seq in REST wild-type and REST KO cells in Fig. 4c and the position of the TF motifs are indicated below.

Extended Data Table 1 | Number of raw and mapped reads and enriched regions for all high-throughput sequencing samples

Type	Sample	Number of raw reads	Number of mapped reads	Percent of mapped reads	Number of enriched regions
RNA-seq	RNA_WT_1	81902703	64369711	79	NA
	RNA_WT_2	73623473	58374618	79	NA
	RNA_WT_3	69703377	55284618	79	NA
	RNA_TKO_1	70027452	55087030	79	NA
	RNA_TKO_2	82516808	64299099	78	NA
	RNA_TKO_3	73579964	58274831	79	NA
	RNA_TKO_CTRL_KD_1	64533537	51328388	80	NA
	RNA_TKO_CTRL_KD_2	79559568	63285076	80	NA
	RNA_TKO_CTRL_KD_3	79175188	62605040	79	NA
	RNA_TKO_NRF1_KD_1	78404841	56795481	72	NA
	RNA_TKO_NRF1_KD_2	74983199	57017015	76	NA
	RNA_TKO_NRF1_KD_3	77279139	57332826	74	NA
DNase-seq	DNASE_WT_1	131089973	96126970	73	125477
	DNASE_WT_2	238165244	170325464	71	222894
	DNASE_TKO_1	210534561	152434201	72	198796
	DNASE_TKO_2	170287886	117016188	69	132369
ChIP-seq	NRF1_CHIP_WT_1	40570927	31414178	77	6835
	NRF1_CHIP_WT_2	40365286	31447763	78	9847
	NRF1_INPUT_WT	22773779	18247061	80	NA
	NRF1_CHIP_TKO_1	32306980	25333581	78	11965
	NRF1_CHIP_TKO_2	45342909	35349643	78	13264
	NRF1_INPUT_TKO	24937026	19810205	79	NA
	NRF1_CHIP_to2i_1	51059626	40940267	80	7088
	NRF1_CHIP_to2i_2	50939344	36209344	71	9470
	NRF1_INPUT_to2i	30416060	23460617	77	NA
	NRF1_CHIP_toSerum_1	42310254	33037271	78	4941
	NRF1_CHIP_toSerum_2	42928737	33018296	77	5562
	NRF1_INPUT_toSerum	25103067	19493583	78	NA
	NRF1_CHIP_Over_1	77223442	56769391	73	18021
	NRF1_CHIP_Over_2	73340571	54380146	74	10479
	NRF1_INPUT_Over	70242507	52149318	74	NA
	NRF1_CHIP_NP_1	117333886	47952332	41	4564
	NRF1_INPUT_NP_1	35321797	15075613	43	NA
	NRF1_CHIP_NP_2	115065799	56350753	49	4906
	NRF1_INPUT_NP_2	25142679	11305749	45	NA
	H3K27ac_CHIP_WT_1	41972346	34720037	83	30616
	H3K27ac_CHIP_WT_2	40822025	34615432	85	29224
	H3K27ac_CHIP_TKO_1	50829570	41308561	81	29455
	H3K27ac_CHIP_TKO_2	45485455	38417647	84	30927
	NRF1_CHIP_HMEC_1	35943107	26822872	75	11585
	NRF1_CHIP_HMEC_2	40718156	28412905	70	13395
	NRF1_INPUT_HMEC	37667963	30523846	81	NA
	NRF1_CHIP_HCC1954_1	41562818	30632702	74	13896
	NRF1_CHIP_HCC1954_2	31412483	22848551	73	12594
	NRF1_INPUT_HCC1954	36664966	29527716	81	NA
Bis-seq	BISSEQ_TKO	257428499	174929585	68	NA
	BISSEQ_to2i	191890338	100546767	52	NA
	BISSEQ_toSerum	215409126	146458573	68	NA

KD = knockdown; CTRL = negative control siRNA; to2i = adapted to 2i (after serum); toSerum = adapted to serum (after 2i); Over = overexpression of NRF1

Exploring the repeat protein universe through computational protein design

TJ Brunette^{1,2*}, Fabio Parmeggiani^{1,2*}, Po-Ssu Huang^{1,2*}, Gira Bhabha³, Damian C. Ekiert⁴, Susan E. Tsutakawa⁵, Greg L. Hura^{5,6}, John A. Tainer^{5,7} & David Baker^{1,2,8}

A central question in protein evolution is the extent to which naturally occurring proteins sample the space of folded structures accessible to the polypeptide chain. Repeat proteins composed of multiple tandem copies of a modular structure unit¹ are widespread in nature and have critical roles in molecular recognition, signalling, and other essential biological processes². Naturally occurring repeat proteins have been re-engineered for molecular recognition and modular scaffolding applications^{3–5}. Here we use computational protein design to investigate the space of folded structures that can be generated by tandem repeating a simple helix–loop–helix–loop structural motif. Eighty-three designs with sequences unrelated to known repeat proteins were experimentally characterized. Of these, 53 are monomeric and stable at 95 °C, and 43 have solution X-ray scattering spectra consistent with the design models. Crystal structures of 15 designs spanning a broad range of curvatures are in close agreement with the design models with root mean square deviations ranging from 0.7 to 2.5 Å. Our results show that existing repeat proteins occupy only a small fraction of the possible repeat protein sequence and structure space and that it is possible to design novel repeat proteins with precisely specified geometries, opening up a wide array of new possibilities for biomolecular engineering.

In repeat proteins, the interactions between adjacent units define the shape and curvature of the overall structure⁶. While in nature the sequences of these units generally differ, stable repeat proteins with identical units have been designed for several families^{7–21} and, for leucine-rich repeats, customized units have allowed for the control of curvature²² and design of new architectures¹⁷. To our knowledge, all designed repeat protein structures to date have been based on naturally occurring families. These families may cover all stable repeat protein structures that can be built from the 20 amino acids or, alternatively, natural evolution may only have sampled a subset of what is possible.

To explore the range of possible repeat protein structures, we generated new repeat protein backbone arrangements and designed sequences predicted to fold into these structures (Fig. 1 and Extended Data Figs 1 and 2). Our designs are entirely *de novo*; they are not based on naturally occurring repeat proteins. The well-packed repeating structures that can be obtained from a simple helix–loop–helix–loop unit are limited to straight rods, and hence we focused on the helix–loop–helix–loop unit from which repeat proteins with a wide diversity of curvatures can be generated. The lengths of the two helices were varied between 10 and 28 residues, and the lengths of the two turns from 1 to 4 residues. Starting conformations for four tandem repeats of each of the 5,776 ($19 \times 19 \times 4 \times 4$) independent combinations of helix and loop lengths were generated by setting the backbone torsion angles

to ideal helix values for helices and extended chain values for loops. Rosetta Monte Carlo fragment assembly²³ was carried out to generate compact structures; each Monte Carlo move was made at the equivalent position in each repeat to preserve symmetry²⁰. Rosetta design calculations²⁴ were then used to identify low-energy amino acid sequences with good core packing²⁵. At each step in the Monte Carlo-simulated annealing design process, a position is picked at random, and the current residue is replaced by a randomly selected amino acid and side-chain conformation (rotamer); a detailed all-atom energy function is then evaluated. Identical substitutions were carried out in each unit to maintain sequence identity between the four repeats; exposed hydrophobic residues in the N- and C-terminal repeats were switched to polar residues in a second round of sequence design to increase solubility. All steps in the design process were completely automated, and the calculations were carried out without manual intervention. Designs with low energies and complementary core side-chain packing were identified, and for the amino acid sequence of each of these designs, multiple independent Rosetta *de novo* folding trajectories²⁶ were carried out starting from an extended chain. The structures and energies of the sampled conformations map out an energy landscape for each protein (Extended Data Fig. 3).

Designed helical repeat proteins (DHRs) for which the design model had much lower energy than any other conformation sampled in the

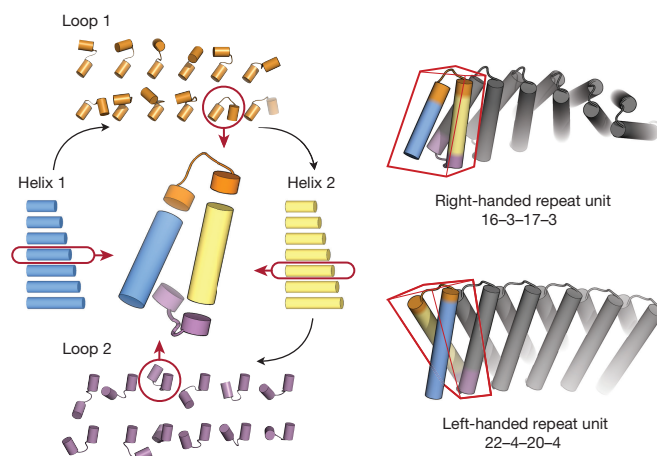


Figure 1 | Schematic overview of the computational design method. Helix–loop–helix–loop combinations are systematically sampled (left) and extended into repeating structures (right) using Rosetta Monte-Carlo fragment assembly. The red boxes on right indicate the individual repeat units; the numbers below, the lengths of helix 1, loop 1, helix 2, and loop 2 for these two examples.

¹Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. ²Institute for Protein Design, University of Washington, Seattle, Washington 98195, USA. ³Department of Cellular and Molecular Pharmacology, UCSF, San Francisco, California 94158, USA. ⁴Department of Microbiology and Immunology, UCSF, San Francisco, California 94158, USA. ⁵Molecular Biophysics & Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. ⁶Department of Chemistry and Biochemistry, University of California, Santa Cruz, California 95064, USA. ⁷Department of Molecular and Cellular Oncology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, USA. ⁸Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA.

*These authors contributed equally to this work.

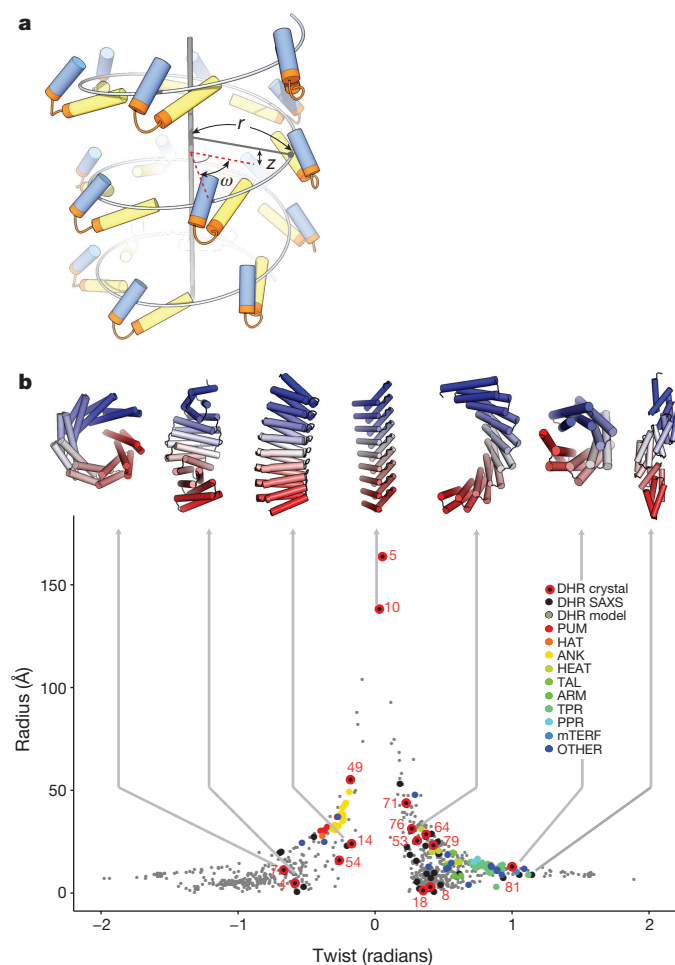


Figure 2 | The helical repeat protein universe. **a**, The geometry of a repeat protein can be described by the radius of the super-helix (r), the axial displacement (z) and the angular displacement or twist (ω) between repeat units. **b**, The 761 DHR models passing all the *in silico* filters are indicated by the small grey circles, the experimentally characterized DHR proteins confirmed by SAXS by large black circles, and those confirmed by X-ray crystallography, by black circles with red borders. The DHR proteins cover radius and twist ranges not found in native repeat protein families (colours). Designs forming right-handed super-helices have positive ω values; left-handed, negative ω values. ANK, ankyrin; ARM, armadillo; TPR, tetratricopeptide repeat; HAT, half TPR; PPR, pentatricopeptide repeat; HEAT, heat repeat; PUM, pumilio homology domain; mTERF, mitochondrial termination factor; TAL, transcription-activator-like effector; OTHER, alpha helical repeat proteins not in the other families. On top, representative experimentally validated designs with a variety of shapes.

de novo folding trajectories were selected and found to span a wide array of architectures. As the rigid body transform relating adjacent repeat units is identical throughout each design by construction, and since the repeated application to an object of an identical rigid body transformation produces a helical array, the designs all have an overall helical structure⁶. It is thus convenient to classify these architectures based on three parameters defining a helix²² (Fig. 2a): the radius (r), the twist between adjacent repeats around the helical axis (ω) and the translation between adjacent repeats along the helical axis (z). Because the repeat units are connected and form well packed structures, the three parameters are coupled. The arc length in the x - y plane spanned by a repeat unit is $\sim r\omega$, and the total length of a unit is $\sim \sqrt{(r\omega)^2 + z^2}$, hence the radius-twist distribution has a hyperbolic shape (Fig. 2b) with highly twisted structures having a smaller radius. Models with high r and high ω do not form a continuous protein core and are discarded during the backbone generation. Similarly, low-energy structures do not have high (>16 Å) z values as helices in adjacent repeats cannot then closely pack (Extended Data Fig. 4). Even with these geometric constraints, the design models span a wide range of helical parameters (Fig. 2b, grey), demonstrating that quite a diversity

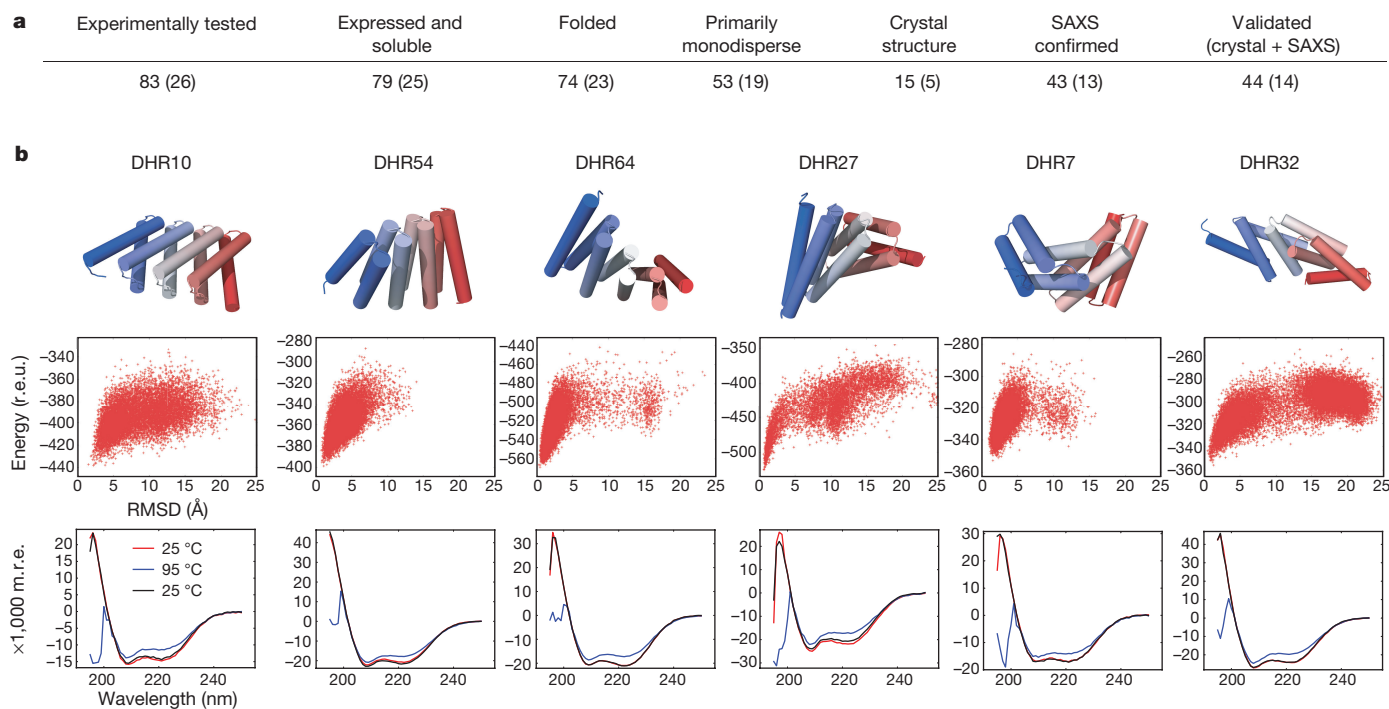


Figure 3 | Characterization of designed repeat proteins. **a**, Design success rate. Values for the subset with disulfide bonds are in parentheses. **b**, Results on six representative designs. Top row, design models. Middle row, computed energy landscapes. Energy is on y axis (r.e.u., Rosetta energy unit) and r.m.s.d. from design model on x axis. All six landscapes

are strongly funnelled into the designed energy minimum. Bottom row, circular dichroism spectra collected at 25 °C (red), 95 °C (blue; high dynode voltages reduce measurement accuracy below 200 nm) and back to 25 °C (black). The proteins do not denature within this temperature range (m.r.e., mean residue ellipticity; deg cm² dmol⁻¹ residue⁻¹).

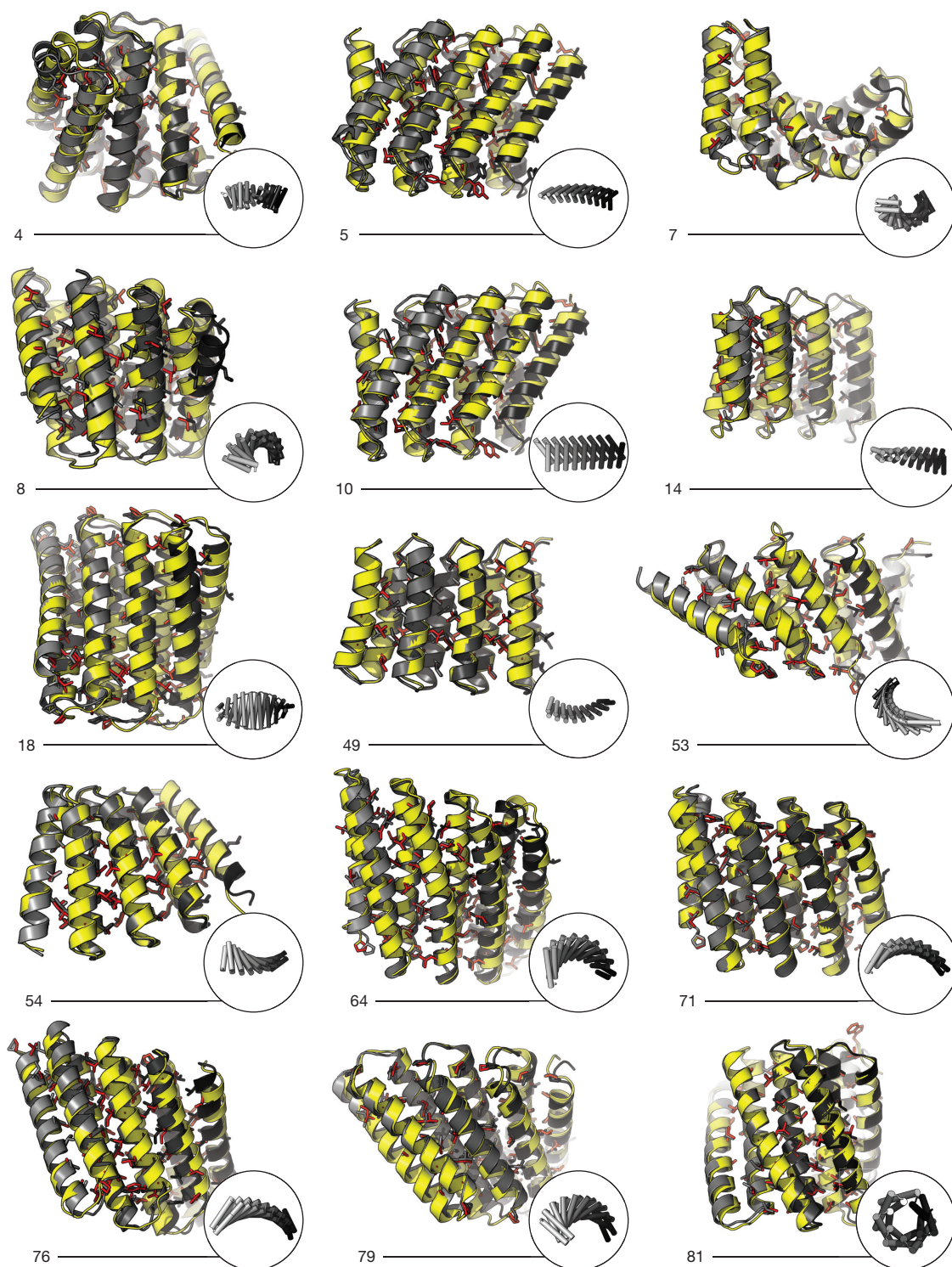


Figure 4 | Crystal structures of 15 designs are in close agreement with the design models. Crystal structures are in yellow, and the design models in grey. Insets in circles show the overall shape of the repeat protein. The r.m.s.d. values across all backbone heavy atoms are: 1.50 Å (DHR4), 1.73 Å (DHR5), 1.30 Å (DHR7), 2.28 Å (DHR8), 1.79 Å (DHR10), 2.38 Å (DHR14), 1.21 Å (DHR18), 0.87 Å (DHR49), 1.33 Å (DHR53), 0.93 Å (DHR54),

1.54 Å (DHR64), 0.67 Å (DHR71), 1.73 Å (DHR76), 1.04 Å (DHR79), 0.65 Å (DHR81). Hydrophobic side chains in the crystal structures (red) are close to those in the designs (grey) (Extended Data Fig. 5). The designed disulfide bonds are formed in the structures of DHR4 and DHR7 but not in the structures of DHR5 and DHR18 due to slight structural shifts relative to the design models.

of structures can be generated by tandem repeating a simple helix–loop–helix–loop unit. In contrast, native helical repeat proteins span a much narrower range of helical parameters (Fig. 2b, colours indicate different families) with very few straight (high r , low ω) or highly twisted (low r , high ω) geometries.

We selected for experimental characterization 83 designs spanning the range of α -helix and loop lengths and overall helical architectures; 26 of these contain disulfide bonds. BLAST searches against the NCBI databases yielded no hits with E values better than 0.0001 for 49 of the designs, and none of the hits found for the remaining designs were to

annotated repeat proteins. HHSEARCH comparisons of the designed repeat units to naturally occurring repeat families in Pfam yielded no hits with an *E* value better than 0.0005 (Supplementary Information Table 4). For each of the designs, we obtained a synthetic gene encoding an amino-terminal capping repeat, two internal repeats, and a carboxy-terminal capping repeat including a six-histidine tag. The proteins were expressed in *Escherichia coli* and purified by affinity chromatography. Of the 83 designs, 74 were expressed in a soluble form and had the expected α -helical circular dichroism spectrum at 25 °C, and 72 were stably folded at 95 °C (Supplementary Experimental Data). Fifty-three of these (64% of the original experimental set) were monomeric by analytical size-exclusion chromatography coupled to multi-angle light scattering; DHR49 and DHR76 were dimeric in solution. Structure stabilization with disulfide bonds did not systematically improve expression, solubility, or folding (Fig. 3a), probably because the designs are already very stable without disulfide bonds. Representative data on six of the designs are shown in Fig. 3b; the data on all 55 proteins is provided as Supplementary Experimental Data.

We solved the crystal structures of 15 of the designs (Fig. 4) with resolutions between 1.20 Å and 3.35 Å. The design models closely match the crystal structures over both the protein backbone (C α root mean square deviations (r.m.s.d.) from 0.7 Å to 2.5 Å) and the hydrophobic core side chains (Fig. 4 and Extended Data Fig. 5; crystal structure and design model side chains are in red and grey respectively). As is evident from the extended models in the Fig. 4 insets, the designs have different overall shapes: for example, DHR10 is linear and untwisted, DHR18 is linear and twisted, DHR8 forms a spiral and DHR81 is a flat toroid. The accuracy of the design models was sufficiently high that all of the crystal structures but DHR5 could be solved by molecular replacement. These repeat proteins are among the largest crystallographically validated protein structures designed completely *de novo*, ranging in size from 171 residues for DHR49 to 238 residues for DHR64. The crystal structures illustrate both the wide range of twist and curvature sampled by our repeat protein generation process and the accuracy with which these can be designed.

To characterize the structures for proteins that were reticent to crystallization and analyse all 55 proteins in solution, we used small-angle X-ray scattering (SAXS)^{27–30}. We collected SAXS profiles for each design, and compared them to scattering profiles calculated from the design models and from crystal structures. For 43 of the designs, the radius of gyration, molecular weight, and distance distributions computed from the SAXS data²⁹ corresponded to those computed from the models (Supplementary Information Table 6). For DHR49 and DHR76, we used the dimer orientation in the crystal for the fitting; the crystallographically confirmed DHR5 was unsuitable for SAXS as it formed higher-order species in solution. To further assess the fit between models and experimental data, we employed the volatility ratio (Vr), which is more robust to experimental noise than the traditional χ^2 comparison used in SAXS³⁰. We used the Vr values of the design models confirmed by crystallography for calibration; designs for which the Vr value between model and experimental data was less than 2.5 were considered successful. All 43 designs with radii, molecular weights, and distances consistent with the SAXS data are below the Vr threshold (Extended Data Fig. 6a). Furthermore, for almost all of the designs, the theoretical scattering profile computed from the design model more closely matches its own experimental scattering profile than the experimental scattering profiles of structurally dissimilar designs (Extended Data Fig. 6b, c).

The crystallographic and SAXS data together structurally validate 44 of the 55 designs that were folded and monodisperse—more than half of the 83 that were experimentally characterized. We randomly selected two designs confirmed by crystallography, two confirmed by SAXS, and two not confirmed by SAXS, and examined their guanidine hydrochloride unfolding profiles. In contrast to almost all native proteins, four of the six designs do not denature at guanidine hydrochloride concentrations up to 7.5 M; the other two, which were confirmed by SAXS but did

not yield crystals, have denaturation midpoints above 3 M (Extended Data Fig. 7). Hence, even the apparent failures are well-folded proteins; small amounts of association may be responsible for the discrepancies between computed and observed SAXS spectra rather than deviations from the design models.

We have shown that a wide range of novel repeat proteins can be generated by tandem repeating a simple helix–loop–helix–loop building block. As illustrated by the comparison of 15 design models to the corresponding crystal structures (Fig. 4), our approach allows precise control over structural details throughout a broad range of geometries and curvatures. The design models and sequences are very different from each other and from naturally occurring repeat proteins, without any significant sequence or structural homology to known proteins (Extended Data Fig. 8). This work achieves key milestones in computational protein design: the design protocol is completely automatic, the folds are unlike those in nature, more than half of the experimentally tested designs have the correct overall structure as assessed by SAXS, and the crystal structures demonstrate precise control over backbone conformation for proteins over 200 amino acids. The observed level of control over the repeating helix–loop–helix–loop architecture shows that computational protein design has matured to the point of providing alternatives to naturally occurring scaffolds, including graded and tunable variation difficult to achieve starting from existing proteins. We anticipate that the 44 successful designs described in this work (Extended Data Fig. 9), and sets generated using similar protocols for other repeat units, will be widely useful starting points for the design of new protein functions and assemblies.

Naturally occurring repeat protein families, such as ankyrins, leucine-rich repeats, TAL effectors and many others, have central roles in biological systems and in current molecular engineering efforts. Our results suggest that these families are only the tip of the iceberg of what is possible for polypeptide chains: there are clearly large regions of repeat protein space that are not sampled by currently known repeat protein structures. Repeat protein structures similar to our designs may not have been characterized yet, or perhaps may simply not exist in nature.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 April; accepted 26 October 2015.

Published online 16 December 2015.

- Kajava, A. V. Tandem repeats in proteins: from sequence to structure. *J. Struct. Biol.* **179**, 279–288 (2012).
- Marcotte, E. M., Pellegrini, M., Yeates, T. O. & Eisenberg, D. A census of protein repeats. *J. Mol. Biol.* **293**, 151–160 (1999).
- Binz, H. K. et al. High-affinity binders selected from designed ankyrin repeat protein libraries. *Nature Biotechnol.* **22**, 575–582 (2004).
- Varadamssetty, G., Tremmel, D., Hansen, S., Parmeggiani, F. & Plückthun, A. Designed Armadillo repeat proteins: library generation, characterization and selection of peptide binders with high specificity. *J. Mol. Biol.* **424**, 68–87 (2012).
- Cortajarena, A. L., Liu, T. Y., Hochstrasser, M. & Regan, L. Designed proteins to modulate cellular networks. *ACS Chem. Biol.* **5**, 545–552 (2010).
- Kobe, B. & Kajava, A. V. When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends Biochem. Sci.* **25**, 509–515 (2000).
- Huang, P.-S., Feldmeier, K., Parmeggiani, F., Fernandez Velasco, D. A., Höcker, B. & Baker, D. *De novo* design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nature Chem. Biol.* <http://dx.doi.org/10.1038/nchembio.1966> (2015).
- Cortajarena, A. L. & Regan, L. Calorimetric study of a series of designed repeat proteins: modular structure and modular folding. *Protein Sci.* **20**, 336–340 (2011).
- Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P. & Plückthun, A. Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.* **332**, 489–503 (2003).
- Mosavi, L. K., Minor, D. L. & Peng, Z. Consensus-derived structural determinants of the ankyrin repeat motif. *Proc. Natl Acad. Sci. USA* **99**, 16029–16034 (2002).

11. Main, E. R. G., Xiong, Y., Cocco, M. J., D'Andrea, L. & Regan, L. Design of stable α -helical arrays from an idealized TPR motif. *Structure* **11**, 497–508 (2003).
12. Urvoas, A. *et al.* Design, production and molecular structure of a new family of artificial alpha-helical repeat proteins (α Rep) based on thermostable HEAT-like repeats. *J. Mol. Biol.* **404**, 307–327 (2010).
13. Lee, S.-C. *et al.* Design of a binding scaffold based on variable lymphocyte receptors of jawless vertebrates by module engineering. *Proc. Natl Acad. Sci. USA* **109**, 3299–3304 (2012).
14. Parmeggiani, F. *et al.* Designed Armadillo repeat proteins as general peptide-binding scaffolds: consensus design and computational optimization of the hydrophobic core. *J. Mol. Biol.* **376**, 1282–1304 (2008).
15. Yadi, I. & Tawfik, D. S. Reconstruction of functional β -propeller lectins via homo-oligomeric assembly of shorter fragments. *J. Mol. Biol.* **365**, 10–17 (2007).
16. Coquille, S. *et al.* An artificial PPR scaffold for programmable RNA recognition. *Nature Commun.* **5**, 5729 (2014).
17. Rämisch, S., Weininger, U., Martinsson, J., Akke, M. & André, I. Computational design of a leucine-rich repeat protein with a predefined geometry. *Proc. Natl Acad. Sci. USA* **111**, 17875–17880 (2014).
18. Lee, J. & Blaber, M. Experimental support for the evolution of symmetric protein architecture from a simple peptide motif. *Proc. Natl Acad. Sci. USA* **108**, 126–130 (2011).
19. Voet, A. R. D. *et al.* Computational design of a self-assembling symmetrical β -propeller protein. *Proc. Natl Acad. Sci. USA* **111**, 15102–15107 (2014).
20. Parmeggiani, F. *et al.* A general computational approach for repeat protein design. *J. Mol. Biol.* **427**, 563–575 (2015).
21. Tripp, K. W. & Barrick, D. Enhancing the stability and folding rate of a repeat protein through the addition of consensus repeats. *J. Mol. Biol.* **365**, 1187–1200 (2007).
22. Park, K. *et al.* Control of repeat-protein curvature by computational protein design. *Nature Struct. Mol. Biol.* **22**, 167–174 (2015).
23. Huang, P.-S. *et al.* RosettaRemodel: a generalized framework for flexible backbone protein design. *PLoS ONE* **6**, e24109 (2011).
24. Leaver-Fay, A. *et al.* ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).
25. Huang, P.-S. *et al.* High thermodynamic stability of parametrically designed helical bundles. *Science* **346**, 481–485 (2014).
26. Bradley, P., Misura, K. M. S. & Baker, D. Toward high-resolution *de novo* structure prediction for small proteins. *Science* **309**, 1868–1871 (2005).
27. Rambo, R. P. & Tainer, J. A. Super-resolution in solution X-ray scattering and its applications to structural systems biology. *Annu. Rev. Biophys.* **42**, 415–441 (2013).
28. Hura, G. L. *et al.* Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nature Methods* **6**, 606–612 (2009).
29. Rambo, R. P. & Tainer, J. A. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* **496**, 477–481 (2013).
30. Hura, G. L. *et al.* Comprehensive macromolecular conformations mapped by quantitative SAXS analyses. *Nature Methods* **10**, 453–454 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Kim and members of the protein production facility at the Institute for Protein Design. This work was facilitated through the use of advanced computational, storage and networking infrastructure provided by the Hyak supercomputer system at the University of Washington. This work was supported in part by grants from the National Science Foundation (NSF) (MCB-1445201 and CHE-1332907), the Defense Threat Reduction Agency (DTRA), the Air Force Office of Scientific Research (AFOSR) (FA950-12-10112) and the Howard Hughes Medical Institute (HHMI-027779). F.P. was the recipient of a Swiss National Science Foundation Postdoc Fellowship (PBZHP3-125470) and a Human Frontier Science Program Long-Term Fellowship (LT000070/2009-L). SAXS work at the Advanced Light Source SIBLYS beamline was supported by the National Institutes of Health grant MINOS (Macromolecular Insights on Nucleic Acids Optimized by Scattering) GM105404 and by United States Department of Energy program Integrated Diffraction Analysis Technologies (IDAT). D.C.E. is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (Grant DRG-2140-12). G.B. is a recipient of the Merck fellowship of the Damon Runyon Cancer Research Foundation (DRG-2136-12) and is supported by NIH grant K99GM112982. J.A.T. is supported by a Robert A. Welch Distinguished Chair in Chemistry. We thank J. Holton for advice on S-SAD data collection, and the staff of ALS 8.2.1 and 8.3.1 for beamline support. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under Contract No. DE-AC02-05CH11231. ALS beamline 8.3.1 is supported by the UC Office of the President, Multicampus Research Programs and Initiatives grant MR-15-338599 and the Program for Breakthrough Biomedical Research, which is partially funded by the Sandler Foundation. ALS beamline 8.2.1 and the Berkeley Center for Structural Biology are supported in part by the National Institutes of Health, National Institute of General Medical Sciences, and the Howard Hughes Medical Institute.

Author Contributions P.-S.H., F.P. and D.B. conceived the *de novo* repeat protein design project. T.B., F.P., P.-S.H. and D.B. conceived the large scale conformational sampling approach. T.B. developed the algorithm with help from F.P. and P.-S.H. F.P. and T.B. expressed and characterized the design with help from P.-S.H. G.B. and D.C.E. setup crystallization trials and solved the crystal structures. F.P., S.E.T., G.L.H., J.T. collected and analysed the SAXS data. F.P., T.B., P.-S.H. and D.B. wrote the manuscript with help from all the authors.

Author Information Crystal structures have been deposited in the RCSB Protein Data Bank with the accession numbers 5CWB (DHR4), 5CWC (DHR5), 5CWD (DHR7), 5CWF (DHR8), 5CWG (DHR10), 5CWH (DHR14), 5CWI (DHR18), 5CWJ (DHR49), 5CWL (DHR53), 5CWL (DHR54), 5CWM (DHR64), 5CWN (DHR71), 5CWO (DHR76), 5CWP (DHR79) and 5CWQ (DHR81). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.B. (dabaker@uw.edu).

METHODS

Code availability. The Rosetta macromolecular modelling suite is available from (<http://www.rosettacommons.org>). The design strategy is described in detail in the Supplementary Information. The Rosetta design code for each step is provided in Supplementary Information section 'Rosetta_examples'.

Similarity search. BLAST^{31,32} and HHSEARCH³³ sequence similarity searches were performed with default settings. HHSEARCH was run on Pfam³⁴. Sequence alignments were depicted using Jalview³⁵. The structural similarity between designs and known helical repeat proteins was assessed by TM-align³⁶ on RepeatsDB³⁷ representative structures.

Protein expression and characterization. Genes were synthesized and cloned in vector pET21 by GenScript (Piscataway). Proteins were expressed in *E. coli* BL21(DE3), induced with 250 μ M isopropyl- β -D-thiogalactopyranoside (IPTG) overnight at 22°C and purified by metal ion affinity chromatography (IMAC) and size-exclusion chromatography (SEC) as described in ref. 20. Cells were lysed by sonication and the clarified lysate was loaded on a NiNTA superflow column (Qiagen). Lysis and washing buffer was Tris 50 mM, pH 8, NaCl 500 mM, imidazole 30 mM, glycerol 5% v/v. Lysozyme (2 mg ml⁻¹), DNaseI (0.2 mg ml⁻¹) and protease inhibitor cocktail (Roche) were added to the lysis buffer before sonication. Proteins were eluted in Tris 50 mM, pH 8, NaCl 500 mM, imidazole 250 mM, glycerol 5% v/v and dialysed overnight either in Tris 20 mM, pH 8, NaCl 150 mM. Protein concentrations were determined using a NanoDrop spectrophotometer (Thermo Scientific). Except as indicated above, enzymes and chemical were purchased from Sigma-Aldrich. Secondary structure content, thermal stability and denaturation in presence of guanidine hydrochloride (GuHCl) were monitored by circular dichroism using an AVIV 420 spectrometer (Aviv Biomedical). Thermal denaturation was followed at 220 nm in Tris 20 mM, 50 mM NaCl, pH 8. Proteins were considered folded if they had the expected α -helical circular dichroism spectrum at 25°C and had either a sharp transition in thermal denaturation or a loss of less than 20% of 220 nm circular dichroism signal at 95°C. Chemical denaturation was monitored in a 1 cm path-length cuvette at 222 nm with protein concentration of 0.05 mg ml⁻¹ in phosphate buffer 25 mM NaCl 50 mM, pH 7. The GuHCl concentration was automatically controlled by a MicroLab titrator (Hamilton). Oligomeric state was assessed by analytical gel filtration coupled to multiple-angle light scattering (AFG-MALS). A Superdex 75 10/300 GL column (or Superdex 200 increase for DHR59, 84, 93) (GE Healthcare) equilibrated in Tris 20 mM, NaCl 150 mM, pH 8 was used on a HPLC LC 1200 Series (Agilent Technologies) connected to a miniDAWN TREOS (Wyatt Technologies). Protein molecular weights were confirmed by mass spectrometry on a LCQ Fleet Ion Trap Mass Spectrometer (Thermo Scientific). Of the 83 designs, 74 were expressed in a soluble form and had the expected α -helical circular dichroism spectrum at 25°C and 72 were stably folded at 95°C. DHR36 has $T_m = 75^\circ\text{C}$ and DHR13 has a broad transition with $T_m = 62^\circ\text{C}$. Fifty-five of these were predominantly monodisperse. DHR49 and 76 were dimeric in solution. SDS-page gels, circular dichroism spectra, thermal denaturation and size-exclusion chromatography profile *ab initio* folding funnel and SAXS data are shown as Supplementary Information for each of the 55 folded and monodisperse proteins.

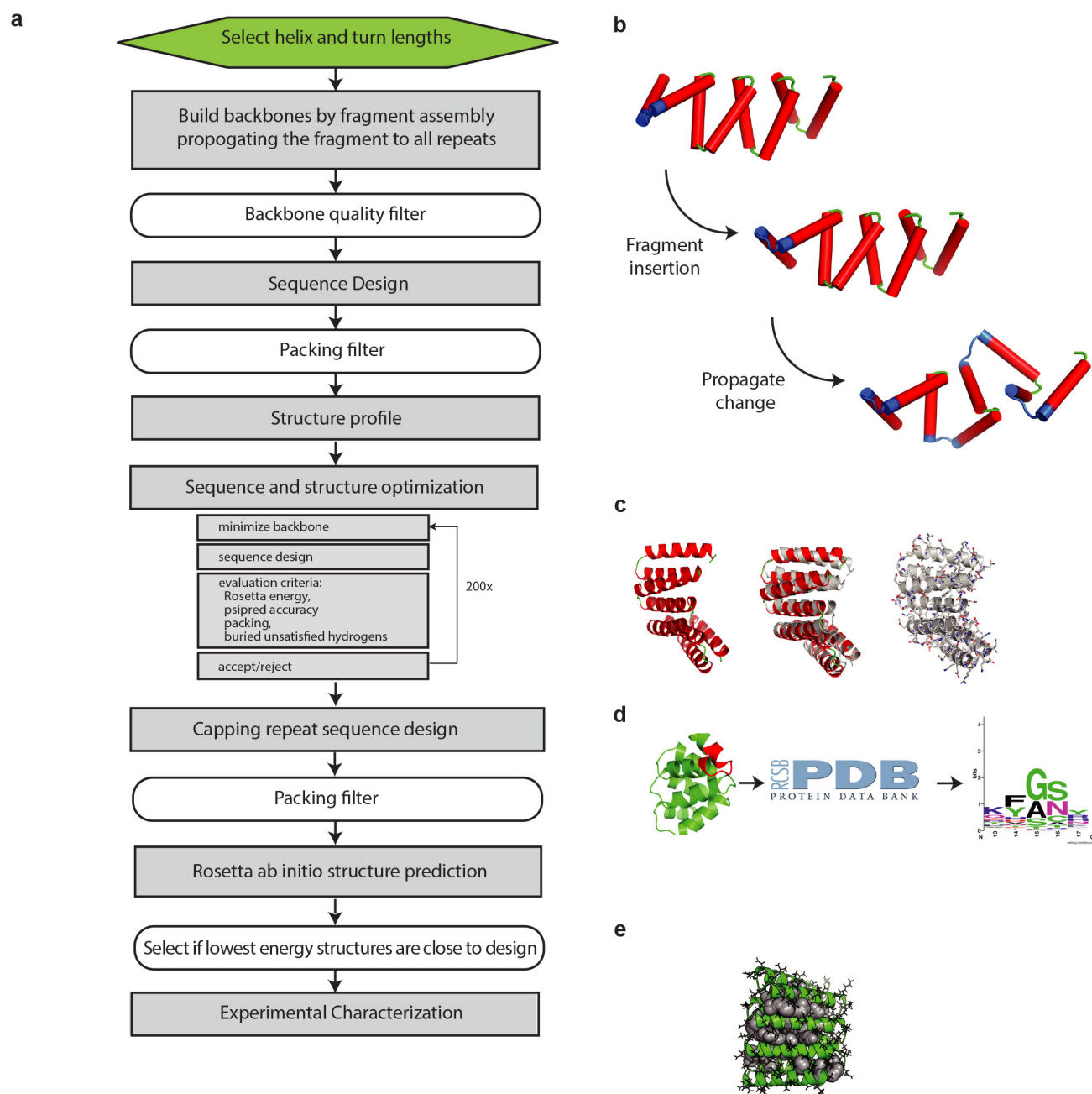
Crystallization. Proteins were purified using NiNTA resin and size-exclusion chromatography on a Superdex 75 column (GE healthcare). Pure fractions in the gel filtration buffer (20 mM Tris pH 8.0, 150 mM NaCl) were pooled and concentrated for crystallography. Final concentrations for each protein are shown in Supplementary Information Table 5. Initial crystallization trials were performed using the JCSG core I-IV screens at 22°C, and crystals were optimized if necessary. Drops were set up with the Mosquito HTS using 100 nl protein and 100 nl of the well solution. Crystals were cryoprotected in the reservoir solution supplemented with ethylene glycol, then flash cooled and stored in liquid nitrogen until data collection. All diffraction data were collected at the Advanced Light Source (ALS) at beamline 8.3.1 or beamline 8.2.1. Crystallization conditions, phasing method and space group information are shown in Supplementary Information Table 5. Data reduction was carried out using XDS³⁸ and HKL2000 (HKL Research). Most of the structures reported here were solved by molecular replacement using Phaser. Search models were generated by *ab initio* folding of the designed sequences in Rosetta and a set of the lowest energy 10–100 models was selected for molecular replacement trials. DHR5 was the only structure which could not be readily solved by molecular replacement. However, owing to the presence of six cysteine residues in the native protein, the DHR5 structure was solved by sulfur single wavelength anomalous dispersion (S-SAD) using a data set collected at 7,235 eV (Supplementary Information). Rigid body, restrained refinement with TLS and simulated annealing were carried out in Phenix³⁹. Manual adjustment of the model was carried out in Coot⁴⁰. The structures were validated using the Quality Control Check v2.8 developed by JCSG, which included Molprobity⁴¹ (publicly available at <http://smb.slac.stanford.edu/jcsg/QC/>). Data collection and final refinement statistics are shown in Supplementary Information Tables 6–14.

SAXS. SAXS data on size-exclusion-chromatography-purified protein were collected at the SIBYLS 12.3.1 beamline at the Advanced Light Source, LBNL^{28,42,43}. Scattering measurements were performed on 20- μ l samples and loaded into a helium-purged sample chamber, 1.5 m from the Mar165 detector. Data were collected on both the original gel filtration fractions and samples concentrated $\sim 2\text{--}8\times$ from individual fractions. Fractions before the void volume and concentrator eluates were used for buffer subtraction. Sequential exposures (0.5, 1, 2, and 5 s) were taken at 12 keV to maximize the signal-to-noise ratio with visual checks for radiation-induced damage to the protein. The data used for fitting were selected for having higher signal to noise ratio and lack of radiation-induced aggregation. In case of concentration dependency, the lowest concentration was used. Models for SAXS comparison were obtained by adding the flexible C-terminal tag present in the constructs to the original designs and the crystal structures, generating 100 trajectories for each starting model by Monte Carlo fragment insertion²³. The results were clustered in Rosetta with a cluster radius of 2 Å and the cluster centres were used for comparison to the experimental data. We used FOXS^{44,45} to calculate scattering profiles from cluster centres and fit them to the experimental data. The quality of fit between models and experimental SAXS data are usually assessed by the χ value⁴⁶, which, however, suffers from over-fitting in case of noisy data sets and domination of the low region of the scattering vector (q) on the value²⁷. To avoid artificially low values that represent false positives, we instead used volatility ratio (V_r)³⁰ as primary metric for fit in the range of $0.015 \text{ \AA}^{-1} < q < 0.25 \text{ \AA}^{-1}$. V_r values of models with available crystal structures range from 0.7 to 2.3 (Supplementary Information Table 15). $V_r = 2.5$ was selected as upper threshold to consider a design as validated by SAXS. An in-depth evaluation of SAXS curves including mass, radius of gyration, Porod number and probability distribution is described in detail in Supplementary Information.

Model profiles for V_r similarity maps were obtained with a standardized fit procedure by averaging the scattering profile of the cluster centres from the five largest clusters and fitting the solvent hydration layer with parameters $C1 = 1.015$ and $C2 = 2.0$ for all the models. V_r was calculated in the range $0.04 \text{ \AA}^{-1} < q < 0.3 \text{ \AA}^{-1}$. The order of display was derived by shape similarity of original computational models using the program damsip⁴⁷ for superposition.

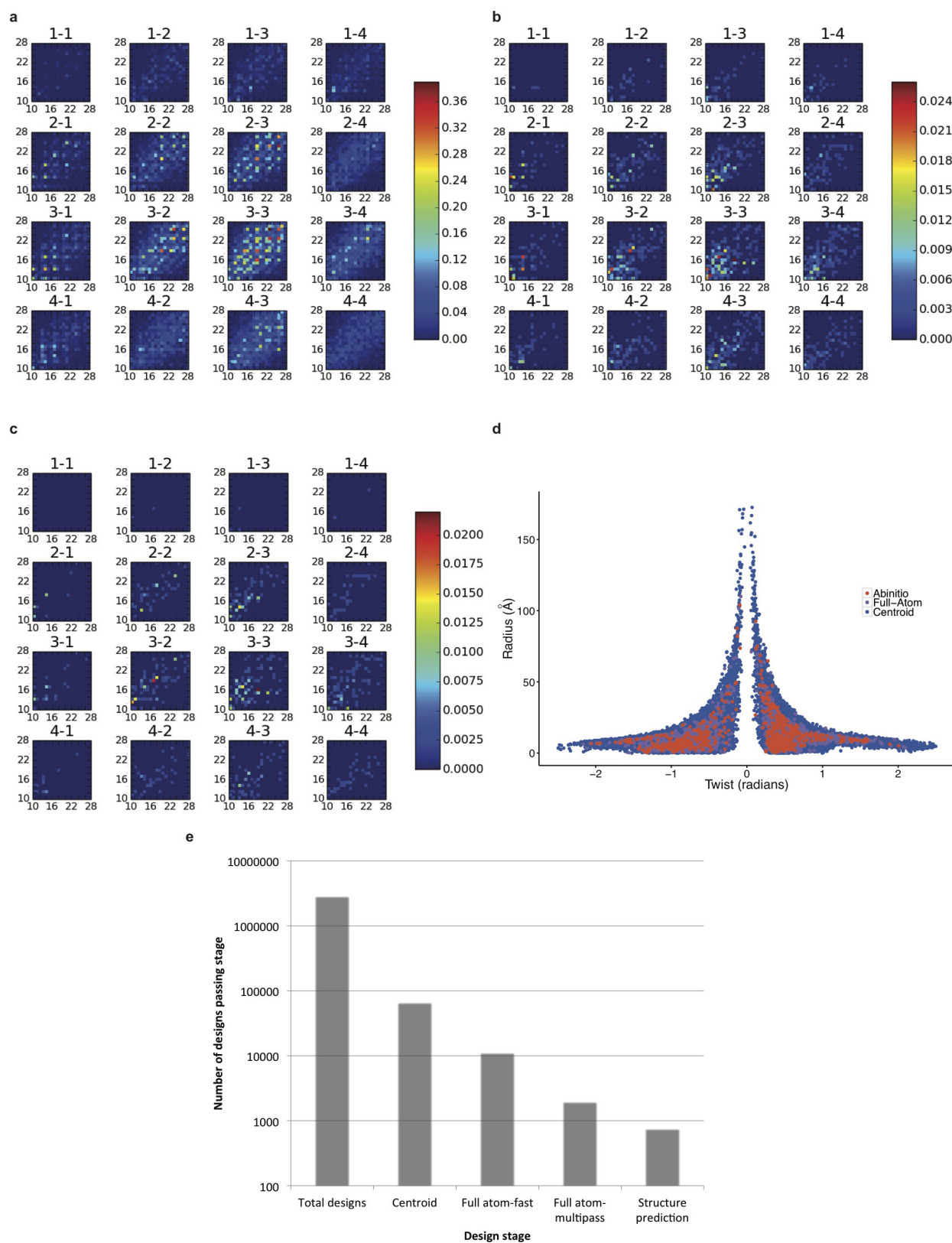
Additional details and discussions on computational design methods, DHR description, experimental characterization, crystallization and SAXS are provided as Supplementary Information.

- Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
- Remmert, M., Biegert, A., Hauser, A. & Söding, J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature Methods* **9**, 173–175 (2012).
- Punta, M. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **40**, D290–D301 (2012).
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
- Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).
- Di Domenico, T. *et al.* RepeatsDB: a database of tandem repeat protein structures. *Nucleic Acids Res.* **42**, D352–D357 (2014).
- Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
- Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
- Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
- Classen, S. *et al.* Implementation and performance of SIBYLS: a dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source. *J. Appl. Crystallogr.* **46**, 1–13 (2013).
- Classen, S. *et al.* Software for the high-throughput collection of SAXS data using an enhanced Blu-Ice/DCS control system. *J. Synchrotron Radiat.* **17**, 774–781 (2010).
- Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys. J.* **105**, 962–974 (2013).
- Schneidman-Duhovny, D., Hammel, M. & Sali, A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res.* **38**, W540–W544 (2010).
- Svergun, D., Barberato, C. & Koch, M. H. J. CRYSOLE – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).
- Petoukhov, M. V. *et al.* New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350 (2012).



Extended Data Figure 1 | Computational protocol for designing *de novo* repeat proteins. **a**, Flowchart of the design protocol. The green box indicates user-controlled inputs, the grey boxes represent steps where protein structure is created or modified, and the white boxes indicate where structures are filtered. **b**, Low-resolution backbone build. **c**, Quick full-atom design (grey) improves the backbone model (red).

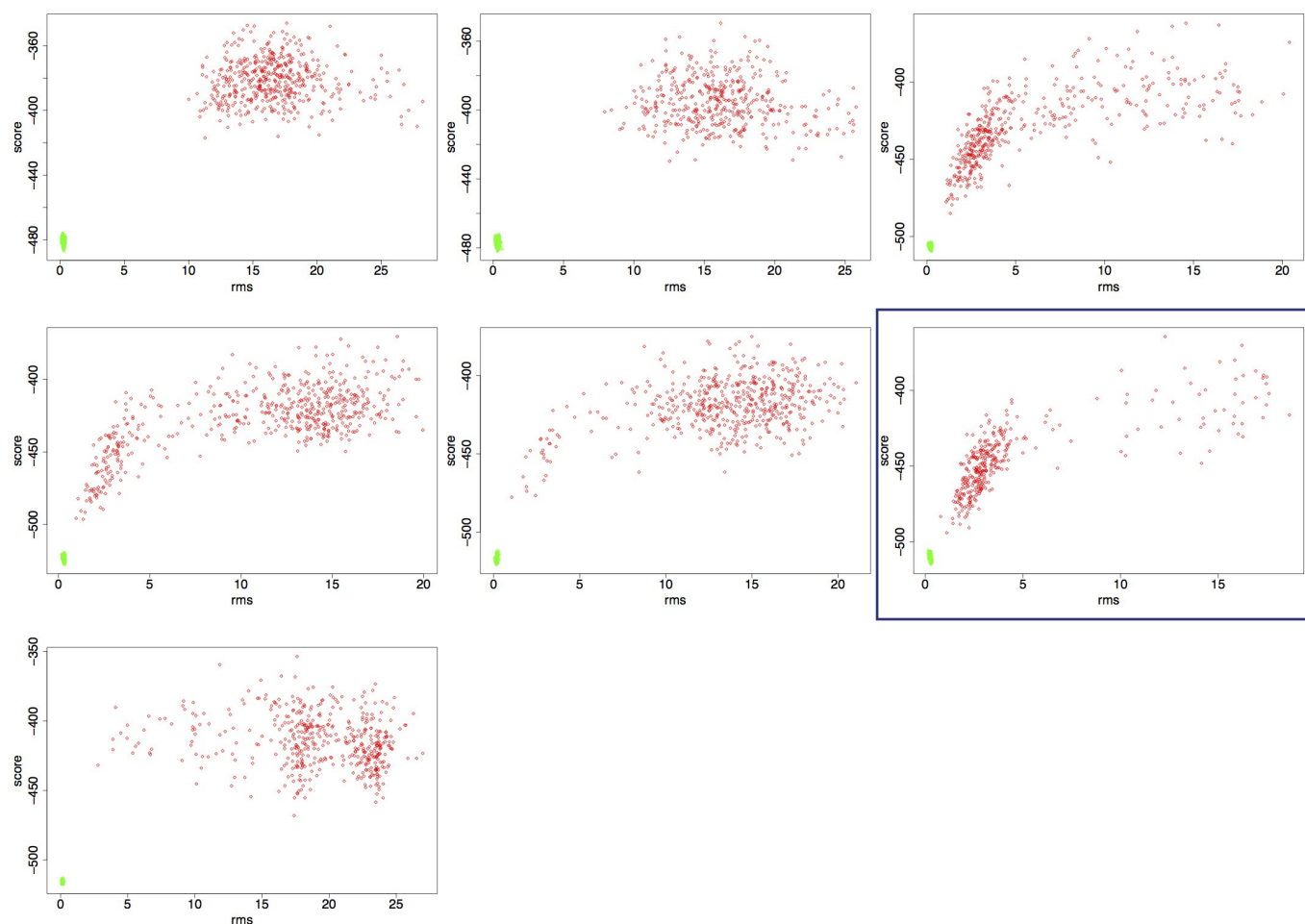
The superposition in the middle highlights the structural changes introduced. **d**, Structural profile: a 9-residue fragment is matched against the Protein Data Bank repository for structures within 0.5 Å r.m.s.d. The sequences from these structures are used to generate a sequence profile that influences design. **e**, Packing filters were used to discard designs with cavities in the core, illustrated as grey spheres.



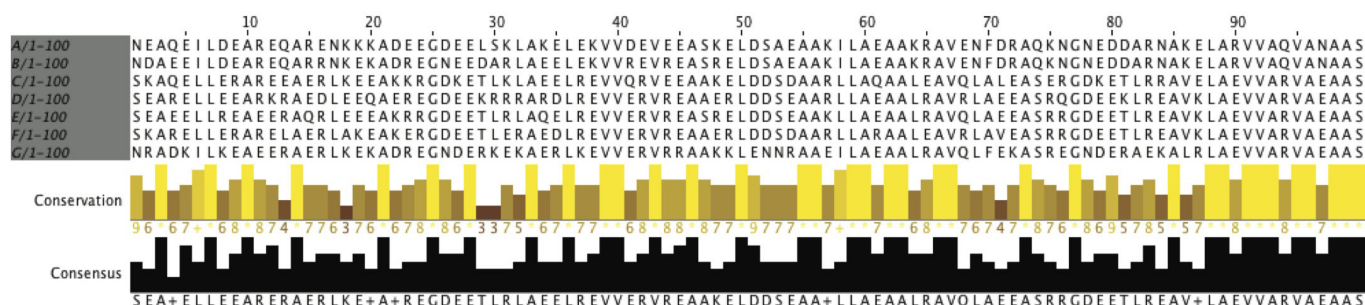
Extended Data Figure 2 | Repeat space explored and model discrimination across design stages. a–c, Percentage of models accepted at backbone building or centroid (a), design (b) and *ab initio* (c) stages. Models are divided according to secondary structure length. The combination of loop 1 and loop 2 lengths is indicated on top. x and y axis indicate helix 1 and helix 2 lengths, respectively. The fraction of models

in the bin that passed the selection stage is indicated in the side bar. Generally, one-residue loops and large differences between helix lengths reduce the number of selected models. d, Distribution of radius and twist of models in the three stages. e, Number of models passing design stages (log scale). From ~2.8 million structures, 761 are accepted.

a-g in order

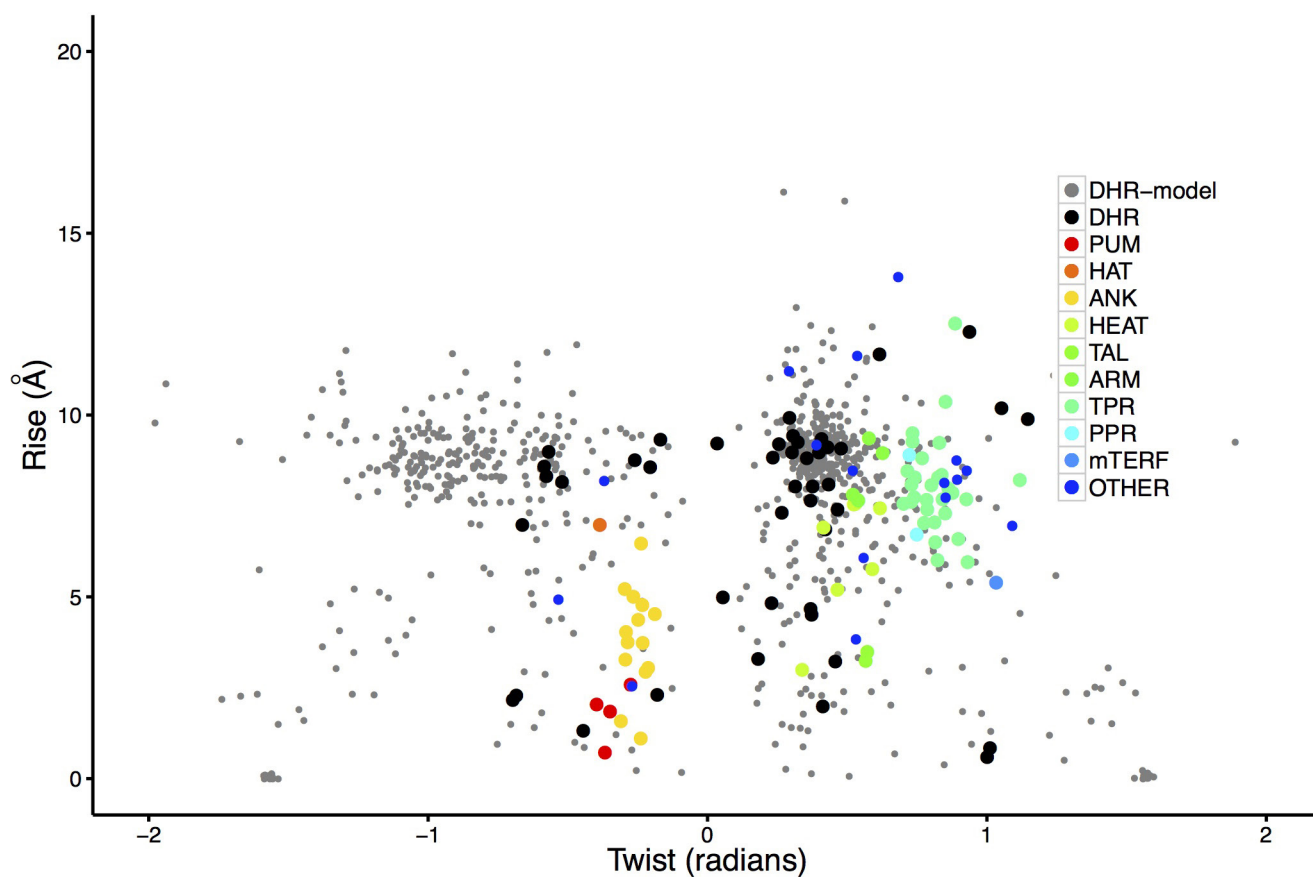


h

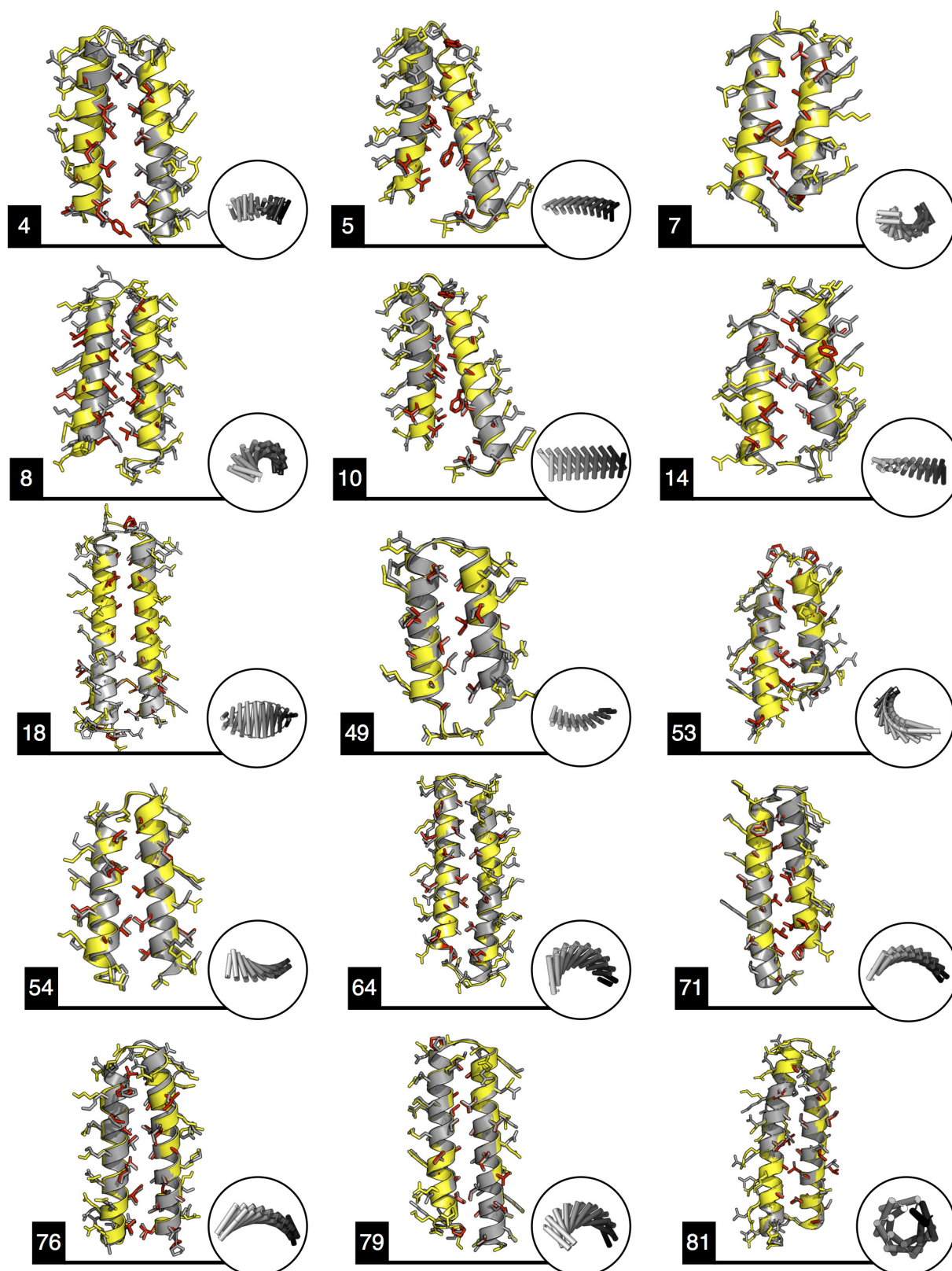


Extended Data Figure 3 | Model validation by *in silico* folding. To assess folding robustness seven sequence variants were made for each design. **a–g**, Illustrate the energy landscape explored by Rosetta *ab initio*. In red are the protein models produced by *ab initio* search, in green by side-chain repacking and minimization (relax). Models in deep global energy minima near the relaxed structures are considered folded.

The variant with highest density of *ab initio* models near the relax region was chosen for experimental characterization (blue box). **h**, Jalview sequence alignment of the first 100 residues of the variants. The yellow bar height indicates sequence conservation, while the black bar represents how often the consensus sequence occurs.

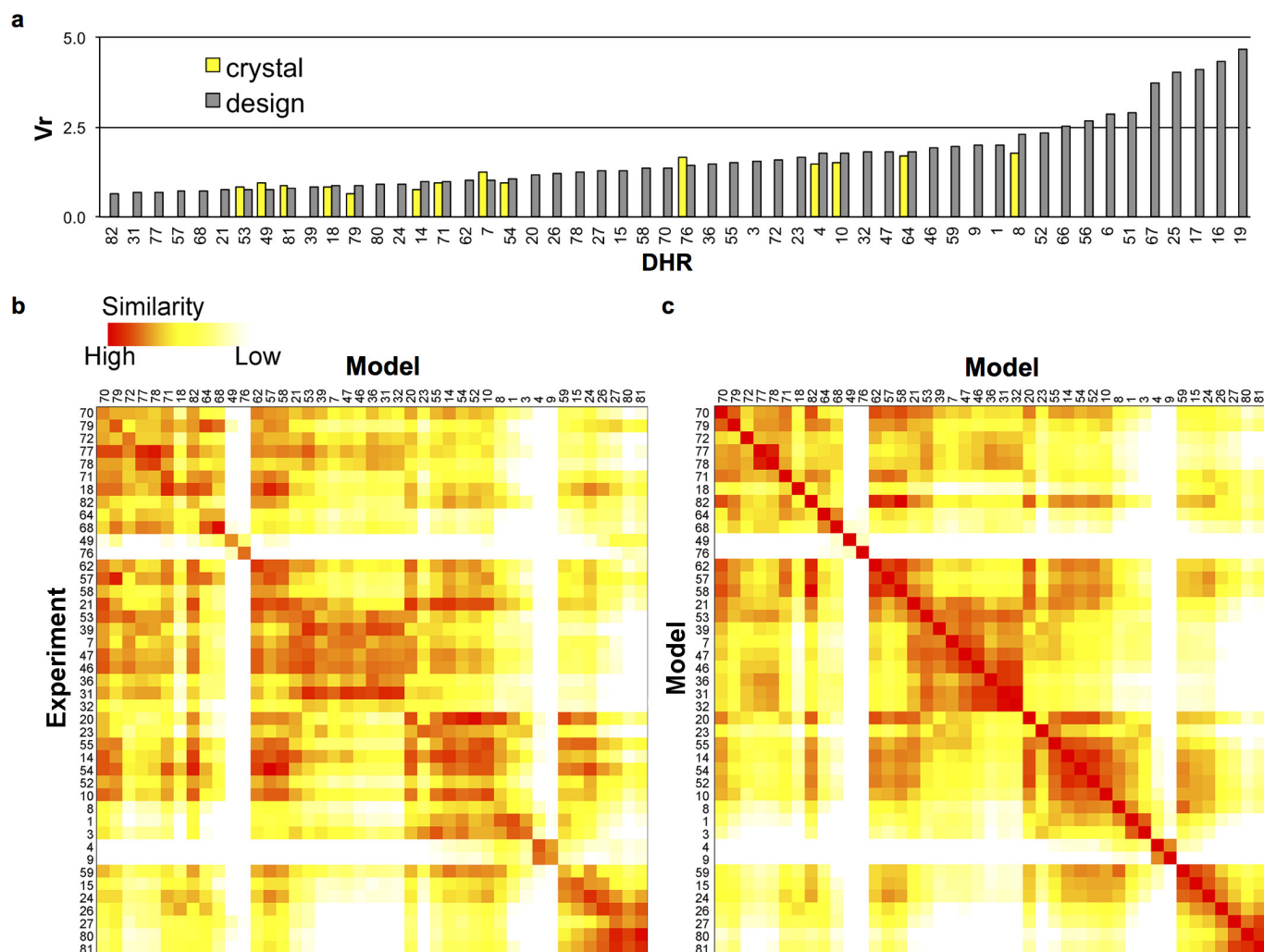


Extended Data Figure 4 | Distribution of DHR axial displacement (z) and twist (ω). Parameters for repeat protein family representatives were extracted as described in the Supplementary Information. The DHR models are the 761 proteins validated by *in silico* folding.



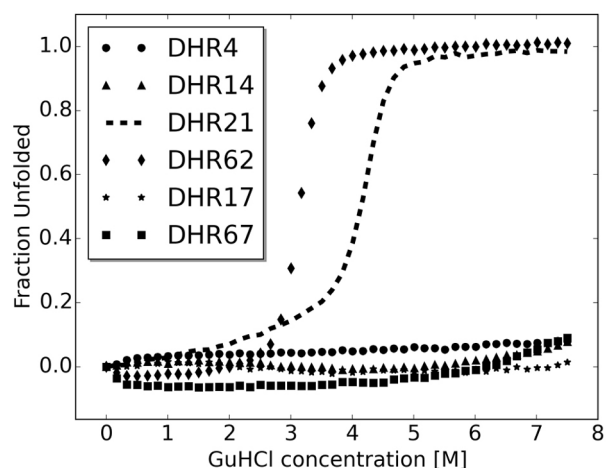
Extended Data Figure 5 | Superposition between single internal repeats (second repeat) of designs (grey) and crystal structures (yellow). Aliphatic and aromatic side chains are in red and cysteines are in orange. DHR7 and 18 show intra repeat disulfide bonds while DHR4 and 81 form

inter-repeat cystines. DHR5 does not form the expected S-S bond. Core side chains in design recapitulate the conformation observed in the crystal structures. Even when the backbone is shifted (for example, DHR5, 8, 15), rotamers are by large correctly predicted.

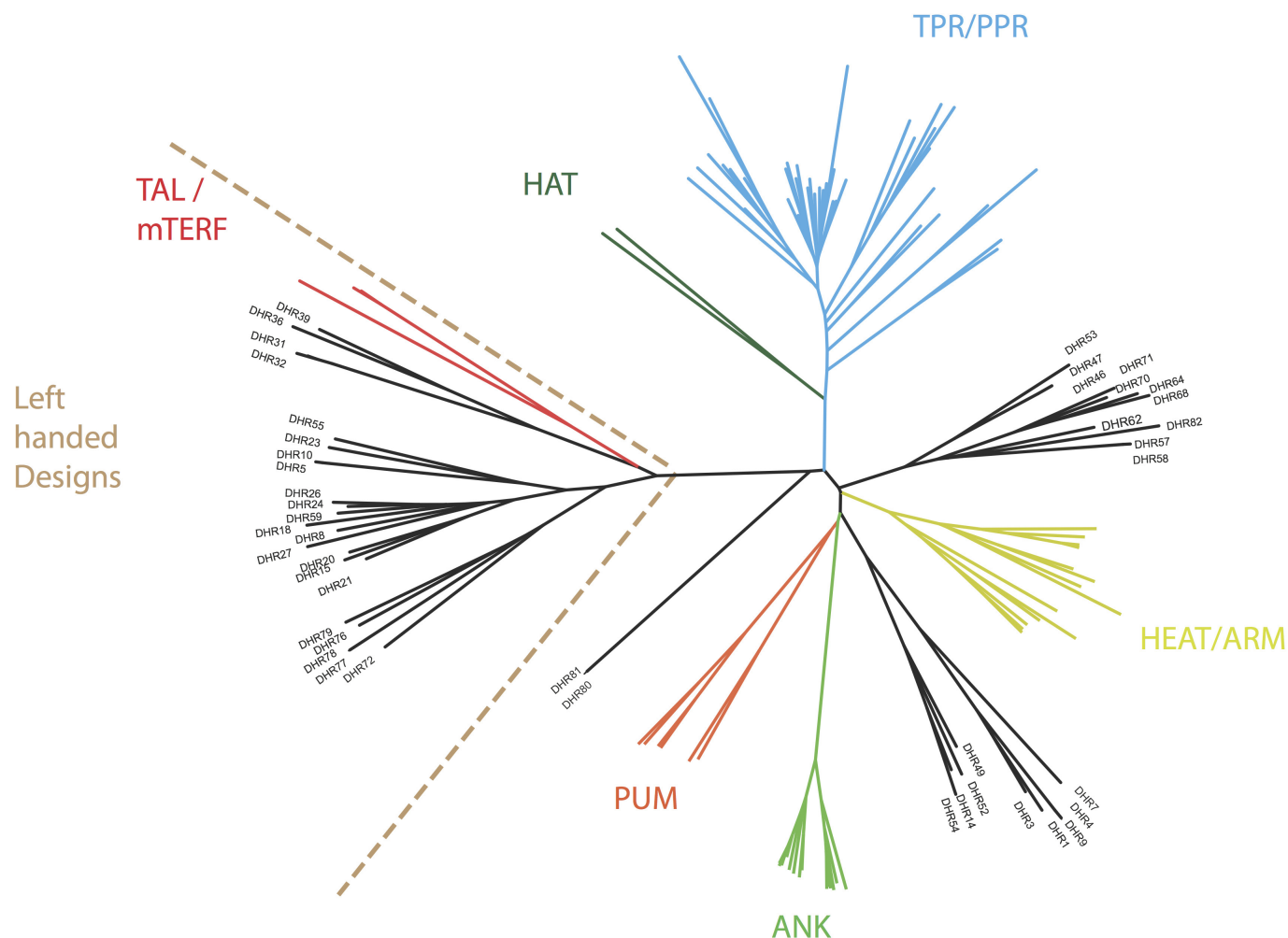


Extended Data Figure 6 | Structural validation by SAXS. **a**, V_r values for the fit of SAXS profiles to design models, in dark grey, and crystal structures, in yellow. For 43 designs, models are within the range defined by crystal structures. DHR49 and DHR76 form dimers in solution and the models employed the configuration observed in the crystal structures. Designs showing aggregation on the scattering profiles, including DHR5 for which the structure was solved, were not included in this figure. **b**, **c**, Pairwise V_r similarity maps³⁰ of 43 design models. **b**, Experimental-to-model profile similarity (**b**) and model-to-model profile similarity (**c**).

Models that are similar to each other show correlation off-diagonal in **c**, and the same pattern is observed when compared to experimental data in **b**. The order of display was obtained by clustering the original designed models by structural similarity. The ability to reproduce characteristic patterns within a large set of designs indicates that the models are capturing the relative structural similarities between proteins in solution. The scores are colour coded with red indicating best agreement and white lack of agreement.

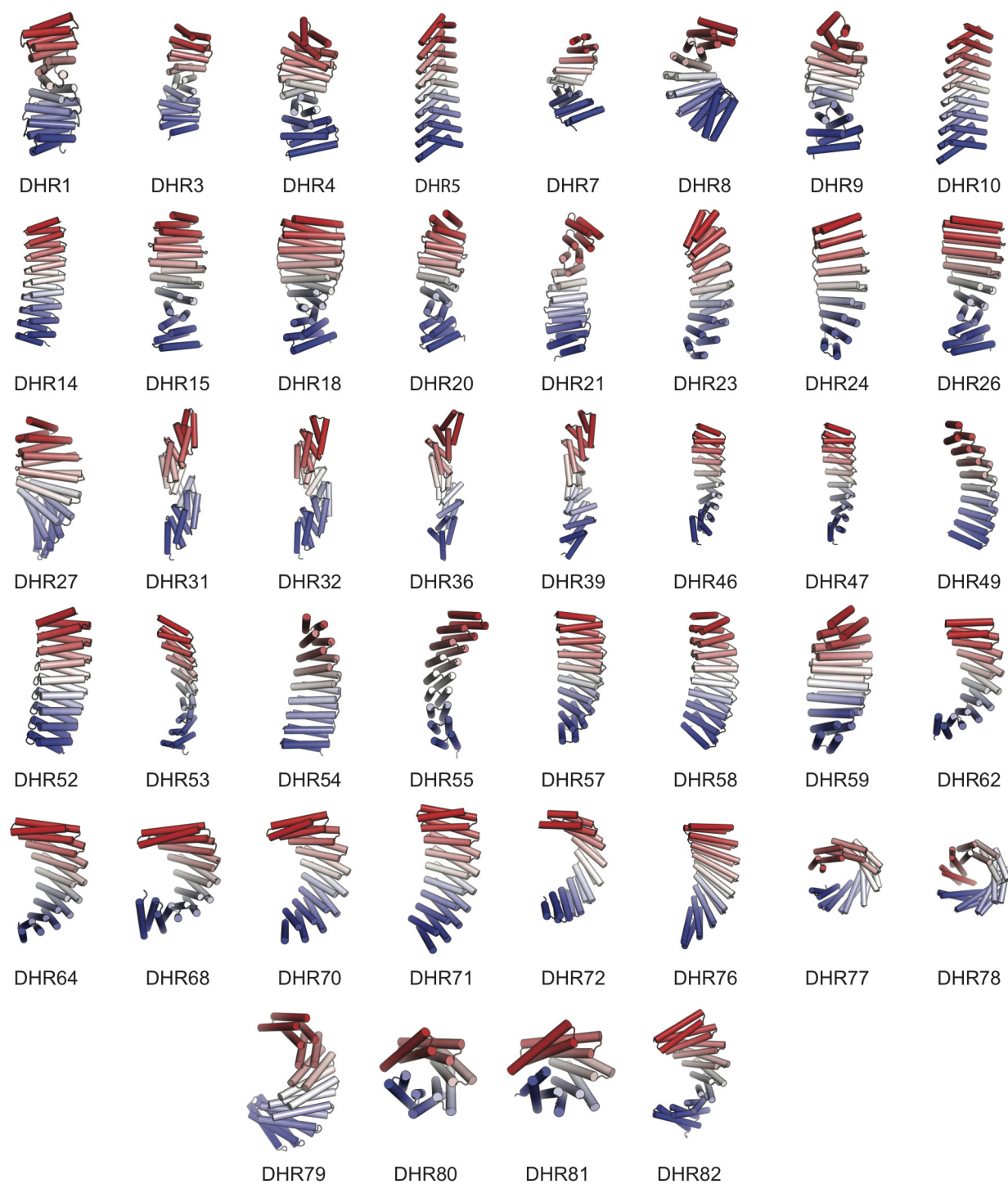


Extended Data Figure 7 | Designs are stable to chemical denaturation by guanidine hydrochloride (GuHCl). Circular-dichroism-monitored GuHCl denaturation experiments were carried for two designs for which crystal structures were solved (DHR4 and DHR14), two with overall shapes confirmed by SAXS (DHR21 and DHR62), and two with overall shapes inconsistent with SAXS (DHR17 and DHR67). In contrast to almost all native proteins, four of the six proteins do not denature at GuHCl concentrations up to 7.5 M. Both designs not confirmed by SAXS were extremely stable to GuHCl denaturation and hence are very well-folded proteins; the discrepancies between the computed and experimental SAXS profiles may be due to small amounts of oligomeric species or variation in overall twist.



Extended Data Figure 8 | Structural similarity between DHRs and repeat protein families. DHRs cluster separately from existing repeat proteins. DHRs are equally distributed between right-handed and left-handed repeats, as referred to the repeat handedness, in contrast to known α -helical repeat proteins, which are mostly right-handed. This result indicates that the handedness observed in known families is not

an intrinsic limitation of repeat proteins structures. Repeat handedness, as defined by Kobe and Kajava⁶, indicates the rotation of the main chain going from the N- to the C-terminal around the axis connecting the repeat centres of mass. The structural similarity tree was built using pairwise comparison as measured by TM-score.



Extended Data Figure 9 | Extended versions of models validated by SAXS and crystallography. DHRs were characterized as containing four repeats but the number of internal repeats can be increased without additional design steps. Extended models highlight the differences in twist and radius between the validated designs.

Rational design of α -helical tandem repeat proteins with closed architectures

Lindsey Doyle¹, Jazmine Hallinan¹, Jill Bolduc¹, Fabio Parmeggiani^{2,3}, David Baker^{2,3,4}, Barry L. Stoddard¹ & Philip Bradley^{1,3,5}

Tandem repeat proteins, which are formed by repetition of modular units of protein sequence and structure, play important biological roles as macromolecular binding and scaffolding domains, enzymes, and building blocks for the assembly of fibrous materials^{1,2}. The modular nature of repeat proteins enables the rapid construction and diversification of extended binding surfaces by duplication and recombination of simple building blocks^{3,4}. The overall architecture of tandem repeat protein structures—which is dictated by the internal geometry and local packing of the repeat building blocks—is highly diverse, ranging from extended, super-helical folds that bind peptide, DNA, and RNA partners^{5–9}, to closed and compact conformations with internal cavities suitable for small molecule binding and catalysis¹⁰. Here we report the development and validation of computational methods for *de novo* design of tandem repeat protein architectures driven purely by geometric criteria defining the inter-repeat geometry, without reference to the sequences and structures of existing repeat protein families. We have applied these methods to design a series of closed α -solenoid¹¹ repeat structures (α -toroids) in which the inter-repeat packing geometry is constrained so as to juxtapose the amino (N) and carboxy (C) termini; several of these designed structures have been validated by X-ray crystallography. Unlike previous approaches to tandem repeat protein engineering^{12–20}, our design procedure does not rely on template sequence or structural information taken from natural repeat proteins and hence can produce structures unlike those seen in nature. As an example, we have successfully designed and validated closed α -solenoid repeats with a left-handed helical architecture that—to our knowledge—is not yet present in the protein structure database²¹.

Engineered proteins that contain closed repeat architectures represent a natural target for rational, geometry-guided design of repeat modules (Fig. 1) for several reasons. Closure results from simple constraints on the inter-repeat geometry: if we consider the transformation between successive repeats as being composed of a rotation (curvature) about an axis together with a translation (rise) parallel to that axis, then the rise must equal zero and the curvature multiplied by the number of repeats must equal a multiple of 360°. Closed structures are stabilized by interactions between the first and last repeats, which obviates the need for capping repeats to maintain solubility and may make them more tolerant to imperfections in the designed geometry than open repeat architectures. Closed repeat arrays offer the advantages of rotational symmetry (for example, in generating higher-order assemblies) with the added control provided by a covalent linkage between subunits. Conversely, it may be possible to convert a monomeric closed repeat protein array into a symmetrical protein assembly by truncation (for example, converting a toroidal protein containing '*n*' repeats into an equivalent homodimeric assembly containing '*n*/2' repeats per subunit) if economy of protein length is required.

We developed an approach to geometry-guided repeat protein design (Fig. 2) that is implemented in the Rosetta molecular modelling package²² and builds on published *de novo* design methodologies²³. Key features include symmetry of backbone and side chain conformations extended across all repeats (allowing computational complexity to scale with repeat length rather than protein length); a pseudo-energy term that favours the desired inter-repeat geometry; clustering and resampling stages that allow intensified exploration of promising topologies; and an *in silico* validation step that assesses sequence–structure compatibility by attempting to re-predict the designed structure given only the designed sequence. Applying this design procedure produced a diverse array of toroidal structures (Fig. 2). We focused primarily on designs with left-handed bundles (Extended Data Fig. 1) since this architecture (closed, left-handed α -solenoid) appears to be absent from the structural database (Supplementary Discussion). We selected five monomeric repeat architectures for experimental characterization: a left-handed 3-repeat family (dTor_3x33L designed toroid with three 33-residue repeats, left-handed), left- and right-handed 6-repeat families (dTor_6x35L and dTor_6x33R), a left-handed 9-repeat family (dTor_9x31L), and a left-handed 12-repeat design built by extending one of the 9-repeat designs by three repeats (dTor_12x31L). To enhance the likelihood of successful expression, purification, and crystallization, we pursued multiple designed sequences for some families, including a round of surface mutants for three designs that were refractory to crystallization (Extended Data Table 1).

We were able to determine five crystal structures for representatives from four monomeric designed toroid families (Fig. 3, Extended Data Fig. 2 and Extended Data Table 2). Close examination of the electron density for the structures, during and after refinement, indicated that most of these highly symmetrical designed proteins display significant rotational averaging within the crystal lattice (Extended Data Fig. 3), such that the positions corresponding to the loops that

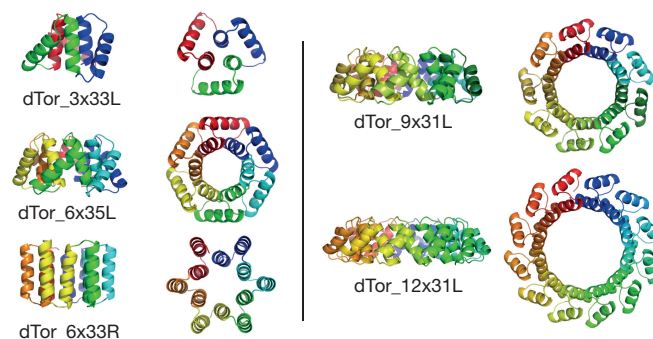


Figure 1 | Designed monomeric repeat architectures. Side and top views of a representative design model from each family are shown in cartoon representation coloured from blue to red as the chain proceeds from the N to the C terminus. Design nomenclature is given in the main text.

¹Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N., Seattle, Washington 98109, USA. ²Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. ³Institute for Protein Design, University of Washington, Seattle, Washington 98195, USA. ⁴Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA. ⁵Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N., Seattle, Washington 98019, USA.

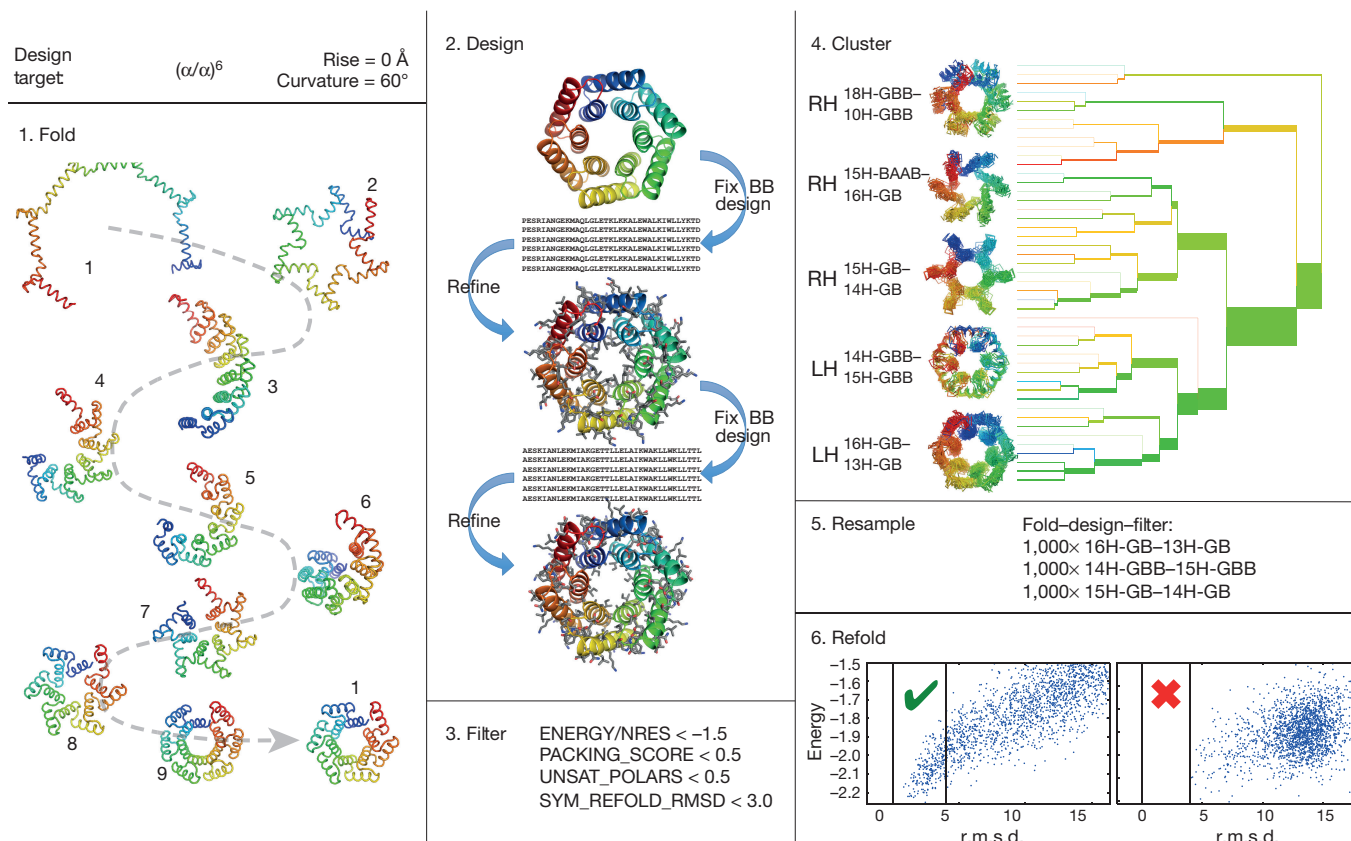


Figure 2 | Overview of the repeat module design process. Given a design target consisting of secondary structure types (α/α in this example), repeat number (6), and desired inter-repeat geometry (rise and curvature), the main steps of the design methodology are (1) symmetric fragment assembly to generate starting backbone conformations; (2) all-atom sequence design and structure relaxation; (3) filtering to eliminate designs with suboptimal per-residue energy (ENERGY/NRES), poor packing (PACKING_SCORE), buried unsatisfied polar atoms (UNSAT_POLARS), or low sequence–structure compatibility (SYM_REFOLD_RMSD, deviation between the final design model and the predicted structure of

the designed sequence; for details see Methods); (4) clustering to identify recurring packing arrangements; (5) intensified sampling of architectures identified in the clustering step; (6) final design assessment by large-scale re-prediction of the designed structure starting from the designed sequence; r.m.s.d., root mean squared deviation. Design cluster identifiers (for example, 14H-GBB–15H-GBB) record the length of the α -helices (14H and 15H) and the backbone conformations of the connecting loops (using a coarse-grained five-state Ramachandran alphabet²⁷; see Methods).

connect each repeated module are occupied by a mixture of continuous peptide and protein termini. This lattice behaviour was observed for most of the structures, but only appeared to significantly affect the refinement *R*-factors for a final multimeric construct (described below) consisting of multiple copies of the first three repeats of dTor_9x31L. In all cases, however, the positions and conformations of secondary structure and individual side chains, which are largely invariant from one repeat to the next, were clear and unambiguous in the respective density maps. Ref. 24 describes similar crystal averaging with associated disorder at protein termini in a set of structures for designed consensus tetratricopeptide repeat (TPR) proteins, albeit with translational averaging along a fibre axis rather than the rotational averaging observed here.

Comparison of the design models with the experimental crystal structures shows that all four designs form left-handed α -helical toroids with the intended geometries. The structural deviation between design model and experimental structure increases with increasing repeat number: from 0.6 Å for the 3-repeat design, to 0.9 Å for the 6-repeat design, to 1.1 Å for the 9- and 12-repeat designs. Inspection of the superpositions in Fig. 3 suggests that the design models are slightly more compact than the experimental structures, a discrepancy which becomes more noticeable as the number of repeats increases. This trend may reflect a tendency of the current design procedure to over-pack side chains during the sequence optimization step (perhaps owing to under-weighting of repulsive electrostatic or van der Waals interactions). Nevertheless, the success of the 12x31L

design implies that, at least for certain repeat modules, it is possible to control the geometry of the central pore by simply varying the number of repeats, without the need to re-optimize the sequence of individual repeats. Further characterization by size-exclusion chromatography indicated that the 3- and 6-repeat designs form stable dimers in solution while the 9- and 12-repeat designs form monomers; all are thermostable (Extended Data Table 1 and Extended Data Figs 4–6). Their behaviour did not vary significantly as a function of protein or salt concentration, nor did they display a dynamic equilibrium between monomeric and dimeric states.

Our ability to successfully design several left-handed α -toroids demonstrates that the apparent absence of this fold from the current database of solved structures is not due to constraints imposed by the helical solenoid architecture or the toroidal geometry. It is possible that there exist in nature left-handed α -toroids whose folds have not been observed; it is also possible that this region of fold space has not been sampled during natural protein evolution. Indeed, left-handed α -helical tandem repeat bundles of any kind—open or closed—are rare relative to their right-handed counterparts (which are found in TPR, Armadillo, HEAT, PUF, and PPR structures, among others). Our search for left-handed helical solenoid repeats with multiple turns in the structural database yielded only the TAL effector^{6,7} and mTERF²⁵ DNA binding domains (Supplementary Discussion). The handedness of our designed toroids is due in part to the use of inter-helical turns whose geometry naturally imparts a handedness to the resulting helical bundle. The three-residue ‘GBB’

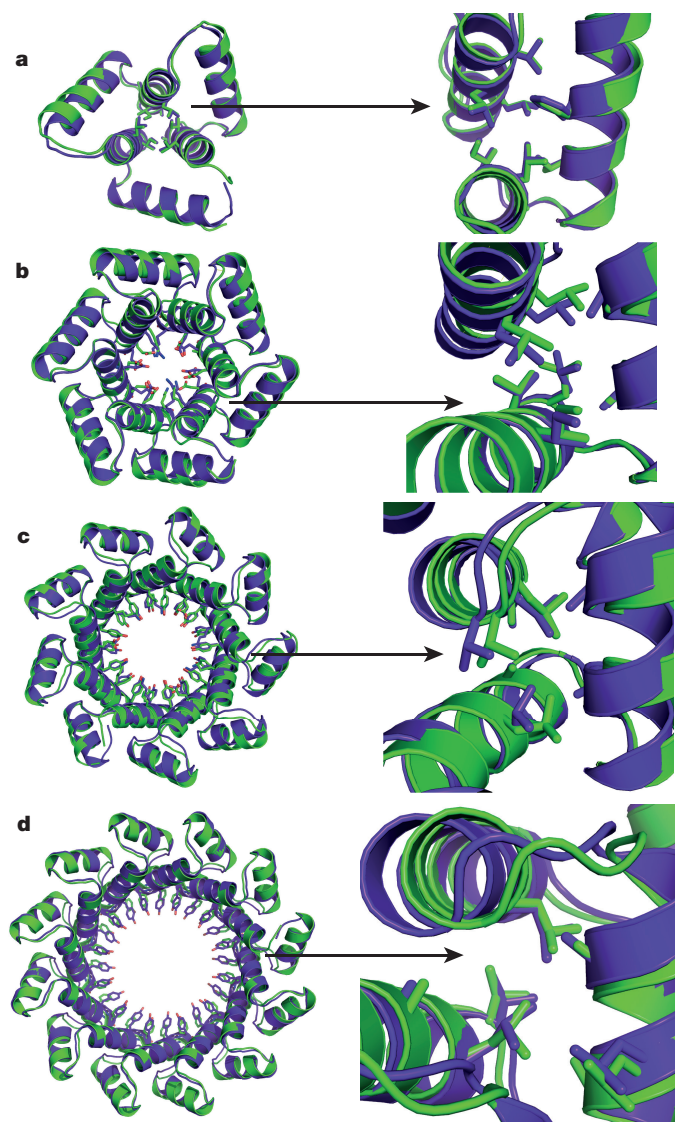


Figure 3 | Superposition of designed toroids (purple) and their refined crystallographic structures (green). Left: the overall superposition of the entire protein backbone, with the side chains that line the innermost pore shown for both models (a, dTor_3x33L; b, dTor_6x35L; c, dTor_9x31L; d, dTor_12x31L). Right: the same superpositions, enlarged to show the packing of side chains and helices between consecutive repeat modules.

(α_L - β - β) turn type used in these designs prefers a left-handed dihelical twist between the connected helices, while the 'GB' turn found in dTor_6x33R correlates with right-handed geometry (Extended Data Fig. 1). Both of these turn types are also compatible with canonical helix capping interactions^{26,27}, which may explain their selection by the design procedure (helix capping guarantees satisfaction of backbone polar groups and also strengthens sequence-encoding of local structure).

We explored the feasibility of splitting one of the larger monomeric designs into fragments that can assemble symmetrically to reform complete toroids comprising multiple copies of identical subunits. We selected the structurally characterized 9x31L design to split into a small 3-repeat subfragment, which was expected to then form a trimeric assembly. This 3-repeat fragment was expressed, purified, and formed diffraction-quality crystals. Upon determination of the experimental structure, we discovered that the design fragment formed an unexpected crystal packing arrangement composed of linked tetrameric rings (that is, containing a total of 12 repeats per ring; Fig. 4a). Indeed, it was this unanticipated finding that led us to synthesize the monomeric

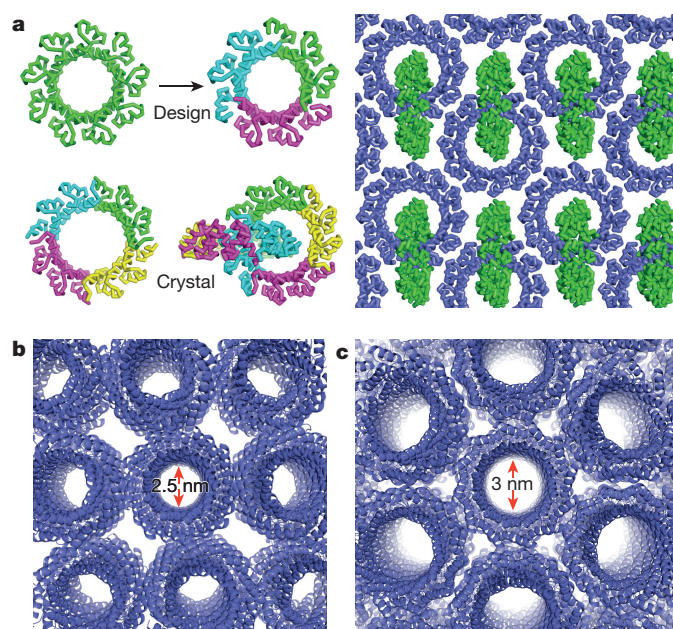


Figure 4 | Crystal packing geometries of designed toroids. a, Rather than forming the expected trimeric toroid ('design'), the 3-repeat subfragment of dTor_9x31L associated in the crystal as two linked tetrameric rings ('crystal') which pack into the layers visualized on the right (the full crystal is then formed from stacks of these layers). Continuous channels are assembled from stacked toroids in the crystals of the monomeric 9x31L and 12x31L designs (b and c respectively).

12x31L design whose characterization demonstrated that the designed 31-residue repeat sequence is compatible with both 9- and 12-repeat monomeric toroidal geometries (and presumably 10- and 11-repeat geometries as well). The crystal structure of the 3-repeat fragment suggests that the 12x geometry may be preferred, and indeed this would be consistent with the apparent tendency of our design procedure to over-pack the design models.

We expect that designed α -toroids may have potential applications as scaffolds for binding and catalysis and as building blocks for higher-order assemblies. Amino acids lining the central pores could be mutated to introduce binding or catalytic functionalities and/or sites of chemical modification. The modular symmetry of monomeric toroids could be exploited to array interaction surfaces with prescribed geometries: a designed interface on the external face of the 12x31L design, for example, could be replicated with two-, three-, four-, or six-fold symmetry by repeating the interfacial mutations throughout the full sequence. Thus monomeric toroids could replace multimeric assemblies as symmetry centres in the assembly of protein cages; by breaking the symmetry of the interaction surfaces it may be possible to create more complex heterotypic assemblies with non-uniform placement of functional sites. Examination of the crystalline arrangements formed by our designed toroids suggests the potential for creating specific one- and two-dimensional assemblies: both the monomeric 9x31L and 12x31L crystals have channels extending continuously through the crystal formed from the pores in vertical stacks of toroids (Fig. 4b, c), with two-dimensional layers of toroids running perpendicular to these stacks. Interface design could be applied to stabilize the crystal contacts seen in the existing structures thereby further stabilizing either the crystalline state or these one- or two-dimensional sub-assemblies^{28,29}. Designed toroids with larger pores that crystallize in a similar manner might form crystal structures with channels capable of hosting guest molecules by covalent linkage or noncovalent binding. Stabilization of the catemeric structure (Fig. 4a) formed by the 3-repeat fragment either by cross-linking or interface design could represent a path towards a variety of novel protein-based materials³⁰.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 April; accepted 30 October 2015.

Published online 16 December 2015.

- Marcotte, E. M., Pellegrini, M., Yeates, T. O. & Eisenberg, D. A census of protein repeats. *J. Mol. Biol.* **293**, 151–160 (1999).
- Kajava, A. V. Tandem repeats in proteins: from sequence to structure. *J. Struct. Biol.* **179**, 279–288 (2012).
- Andrade, M. A., Perez-Iratxeta, C. & Ponting, C. P. Protein repeats: structures, functions, and evolution. *J. Struct. Biol.* **134**, 117–131 (2001).
- Grove, T. Z., Cortajarena, A. L. & Regan, L. Ligand binding by repeat proteins: natural and designed. *Curr. Opin. Struct. Biol.* **18**, 507–515 (2008).
- Wang, X., McLachlan, J., Zamore, P. D. & Hall, T. M. Modular recognition of RNA by a human pumilio-homology domain. *Cell* **110**, 501–512 (2002).
- Mak, A. N., Bradley, P., Cernadas, R. A., Bogdanove, A. J. & Stoddard, B. L. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* **335**, 716–719 (2012).
- Deng, D. *et al.* Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* **335**, 720–723 (2012).
- Barkan, A. *et al.* A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet.* **8**, e1002910 (2012).
- Reichen, C., Hansen, S. & Plückthun, A. Modular peptide binding: from a comparison of natural binders to designed armadillo repeat proteins. *J. Struct. Biol.* **185**, 147–162 (2014).
- Wierenga, R. K. The TIM-barrel fold: a versatile framework for efficient enzymes. *FEBS Lett.* **492**, 193–198 (2001).
- Kobe, B. & Kajava, A. V. When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends Biochem. Sci.* **25**, 509–515 (2000).
- Main, E. R., Xiong, Y., Cocco, M. J., D'Andrea, L. & Regan, L. Design of stable alpha-helical arrays from an idealized TPR motif. *Structure* **11**, 497–508 (2003).
- Binz, H. K. *et al.* High-affinity binders selected from designed ankyrin repeat protein libraries. *Nature Biotechnol.* **22**, 575–582 (2004).
- Parmeggiani, F. *et al.* Designed armadillo repeat proteins as general peptide-binding scaffolds: consensus design and computational optimization of the hydrophobic core. *J. Mol. Biol.* **376**, 1282–1304 (2008).
- Urvoas, A. *et al.* Design, production and molecular structure of a new family of artificial alpha-helical repeat proteins (α Rep) based on thermostable HEAT-like repeats. *J. Mol. Biol.* **404**, 307–327 (2010).
- Boersma, Y. L. & Plückthun, A. DARPins and other repeat protein scaffolds: advances in engineering and applications. *Curr. Opin. Biotechnol.* **22**, 849–857 (2011).
- Rämisch, S., Weininger, U., Martinsson, J., Akke, M. & André, I. Computational design of a leucine-rich repeat protein with a predefined geometry. *Proc. Natl Acad. Sci. USA* **111**, 17875–17880 (2014).
- Voet, A. R. *et al.* Computational design of a self-assembling symmetrical β -propeller protein. *Proc. Natl Acad. Sci. USA* **111**, 15102–15107 (2014).
- Park, K. *et al.* Control of repeat-protein curvature by computational protein design. *Nature Struct. Mol. Biol.* **22**, 167–174 (2015).
- Parmeggiani, F. *et al.* A general computational approach for repeat protein design. *J. Mol. Biol.* **427**, 563–575 (2015).
- Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
- Leaver-Fay, A. *et al.* ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).
- Koga, N. *et al.* Principles for designing ideal protein structures. *Nature* **491**, 222–227 (2012).
- Kajander, T., Cortajarena, A. L., Mochrie, S. & Regan, L. Structure and stability of designed TPR protein superhelices: unusual crystal packing and implications for natural TPR proteins. *Acta Crystallogr. D* **63**, 800–811 (2007).
- Jiménez-Menéndez, N. *et al.* Human mitochondrial mTERF wraps around DNA through a left-handed superhelical tandem repeat. *Nature Struct. Mol. Biol.* **17**, 891–893 (2010).
- Aurora, R. & Rose, G. D. Helix capping. *Protein Sci.* **7**, 21–38 (1998).
- Wintjens, R. T., Rooman, M. J. & Wodak, S. J. Automatic classification and analysis of alpha alpha-turn motifs in proteins. *J. Mol. Biol.* **255**, 235–253 (1996).
- Grove, T. Z., Regan, L. & Cortajarena, A. L. Nanostructured functional films from engineered repeat proteins. *J. R. Soc. Interface* **10**, <http://dx.doi.org/10.1098/rsif.2013.0051> (2013).
- Lanci, C. J. *et al.* Computational design of a protein crystal. *Proc. Natl Acad. Sci. USA* **109**, 7304–7309 (2012).
- Abe, S. & Ueno, T. Design of protein crystals in the development of solid biomaterials. *RSC Adv.* **5**, 21366–21375 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors thank Scientific Computing at the Fred Hutchinson Cancer Research Center for providing the computational infrastructure necessary for this project. This research was supported by the following research grants: National Institutes of Health R21GM106117 to P.B. and R01GM49857 to B.L.S., and Swiss National Science Foundation Postdoc Fellowship PBZHP3-125470 and Human Frontier Science Program Long-Term Fellowship LT000070/2009-L to F.P.

Author Contributions L.D., J.H., J.B. and F.P. expressed, purified, and characterized designed constructs. L.D., J.H. and J.B. performed crystal screening, collected diffraction data, and solved crystal structures. P.B. developed and implemented the repeat design algorithms. P.B. performed sequence design calculations with feedback from F.P., B.L.S. and D.B. supervised the research. P.B. conceived of the toroid design project with input from B.L.S. and D.B. P.B. wrote the manuscript with input from the other authors.

Author Information Crystal structures determined in this study have been deposited in the RCSB Protein Data Bank under accession numbers 4YXX (dTor_6x35L), 4YY2 (dTor_3x33L_2-2a), 4YY5 (dTor_3x33L_2-2b), 4YXY (dTor_9x31L_sub), 4YXZ (dTor_9x31L), and 5BYO (dTor_12x31L). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.B. (pbradley@fredhutch.org).

METHODS

Computational design. The repeat module design process applied here consisted of an initial diversification round of large-scale sampling followed by filtering and clustering and then a second intensification round of sampling focused on successful topologies identified in the first round.

Fragment assembly. Starting backbone models for sequence design were built using a fragment assembly protocol which is based on the standard Rosetta *ab initio* protocol³¹ with the following modifications: (1) fragment replacement moves were performed symmetrically across all repeats, guaranteeing that backbone torsion angles were identical at corresponding positions across repeats; (2) a pseudo-energy term (equal to the deviation between actual and desired curvature, in degrees, plus the deviation in rise multiplied by a factor of 5) was added to the potential to favour satisfaction of the geometric constraints; (3) the amino-acid sequence used for low-resolution scoring was assigned randomly at the start of each simulation from secondary-structure-specific distributions (helix: Ala+Ile+Leu+Asp+Ser; turn: Gly+Ser), which had the effect of increasing the diversity in helix packing distances and geometries compared with using a constant sequence such as poly-Val or poly-Leu. At the start of each independent design trajectory, the lengths of the secondary structure elements and turns were chosen randomly, defining the target secondary structure of the repeat module and its length. Together with the number of repeats, this defined the total length of the protein and the complete secondary structure, which was used to select 3- and 9-residue backbone fragments for use in the low-resolution fragment assembly phase. The design calculations reported here sampled helix lengths from 7 to 20 residues, turn lengths from 1 to 5 residues, and total repeat lengths ranging from 20 to 40 residues.

Sequence design. The low-resolution fragment assembly simulation was followed by an all-atom sequence design stage consisting of two cycles alternating between fixed-backbone sequence design and fixed-sequence structure relaxation. Symmetry of backbone and side-chain torsion angles and sequence identities was maintained across all repeats. Since the starting backbones for design were built by relatively coarse sampling in a low-resolution potential, sequences designed with the standard all-atom potential were dominated by small amino acids and the resulting structures tended to be under-packed. To correct for this tendency, a softened Lennard-Jones potential³² was used for the sequence design steps, while the standard potential was used during the relaxation step. The Rosetta *score12p-rime* weights set was used as the standard potential for these design calculations.

Filtering and clustering. Final design models (typically 10,000–100,000 in this study) were first sorted by per-residue energy (total energy divided by the number of residues, to account for varying repeat length) and the top 20% filtered for packing quality (sasapack_score < 0.5), satisfaction of buried polar groups (buried unsatisfied donors per repeat < 1.5, buried unsatisfied acceptors per repeat < 0.5), and sequence-structure compatibility via a fast, low-resolution symmetric refolding test (40 trajectories, requiring at least 1 under an r.m.s.d. threshold of 2 Å for 3-repeat designs and 4 Å for larger designs). Designs that passed these filters were clustered by C- α r.m.s.d. (allowing for register shifts when aligning helices with unequal lengths) to identify recurring architectures. The clusters were ranked by averaging residue energy, packing quality, and refolding success over all cluster members.

Resampling. During the intensification round of designs, representative topologies from successful design clusters were specifically resampled by enforcing their helix and turn lengths as well as their turn conformations (defined using a five-state, coarse-grained backbone torsion alphabet²⁷; Extended Data Fig. 1e) during fragment selection.

Large-scale refolding. Selected low-energy designs from the second round that pass the filters described above were evaluated by a large-scale refolding test in which 2,000–10,000 *ab initio* models were built by standard (asymmetric) fragment assembly followed by all-atom relaxation. Success was measured by assessing the fraction of low-energy *ab initio* models with r.m.s.d. values to the design model under a length-dependent threshold.

Symmetry-breaking in the central pore. For designed toroids with an open, polar central pore, perfect symmetry may not allow optimal electrostatic interactions between nearby side chains corresponding to the same repeat position in successive repeats. We therefore explored symmetry-breaking mutations at a handful of inward-pointing positions via fixed-backbone sequence design simulations in which the length of the repeating sequence unit was doubled/tripled (for example, whereas perfect six-fold repeat symmetry would require K-K-K-K-K-K or E-E-E-E-E-E, doubling the repeat length allows charge complementarity with K-E-K-E-K-E). Solutions from these designs were accepted if they significantly lowered the total energy.

Design model for dTor_12x31L. The 12x31L design construct was generated by duplicating the final three repeats of the 9x31L design. To build a 'design model' for comparison with the experimentally determined structure, we followed the

resampling protocol now forcing the 12x31L amino-acid sequence in addition to the number of repeats (12) and the helix and turn lengths (H14-L3-H11-L3) and turn conformations (GBB). Thus the sequence design steps were reduced to rotamer optimization (since the amino-acid identities were fixed). This symmetric structure prediction process was repeated 10,000 times and the lowest-energy final model was taken as the computational model.

Surface mutations to enhance crystallization. For a single representative of the 3x31L and 6x31R families, we performed lattice docking and design simulations to select mutations that might promote crystallization. Core positions were frozen at the design sequence. Candidate space groups were selected from those most commonly observed in the protein structural database. Theoretical models of crystal packing arrangements were built by randomly orienting the design model within the unit cell and reducing the lattice dimensions until clashes were encountered. Symmetric interface design was performed on these docked arrangements, and final designs were filtered by energy, packing, satisfaction of polar groups, and number of mutations from the original design model.

Handedness of tandem repeat helical bundles. To compute the handedness of helical bundles formed by tandem repeat proteins, we generated an approximate helical bundle axis curve by joining the location of repeat-unit centres of mass in a sliding fashion along the protein chain. The handedness was then estimated by computing the directionality of the winding of the polypeptide chain about this axis curve.

Structural bioinformatics. To assess similarity between design models and proteins in the structural database, we performed searches using the structure-structure comparison program DALI³³ as well as consulting the protein structure classification databases CATH³⁴, SCOPe³⁵, and ECOD³⁶. Further details are given in Supplementary Discussion.

Code availability. Repeat protein design methods were implemented in the Rosetta software suite (www.rosettacommons.org) and will be made freely available to academic users; licenses for commercial use are available through the University of Washington Technology Transfer office.

Cloning and protein expression. The plasmids encoding individual constructs were cloned into previously described bacterial pET15HE expression vectors³⁷ containing a cleavable N-terminal His-tag and an ampicillin resistance cassette. Sequence-verified plasmids were transformed into BL21(DE3)RIL *Escherichia coli* cells (Agilent Technologies) and plated on lysogeny broth (LB) medium with ampicillin (100 μ g ml⁻¹). Colonies were individually picked and transferred to individual 10 ml aliquots of LB-ampicillin media and shaken overnight at 37 °C. Individual 10 ml aliquots of overnight cell cultures were added to individual 1 l volumes of LB-ampicillin, which were then shaken at 37 °C until the cells reached an absorbance at 600 nm of 0.6–0.8. The cells were chilled for 20 min at 4 °C, then isopropyl- β -D-thiogalactoside (IPTG) was then added to each flask to a final concentration of 0.5 mM to induce protein expression. The flasks were shaken overnight at 16 °C, and then pelleted by centrifugation and stored at –20 °C until purification.

Construct dTor_6x35L(SeMet), incorporating a single methionine residue at position 168 in the original design construct, was generated using a QuikChange site-directed mutagenesis kit (Agilent) and corresponding protocol from the vendor. The resulting plasmid construct was again transformed into BL21(DE3)RIL *E. coli* cells (Agilent Technologies) and plated on LB plates containing ampicillin (100 μ g ml⁻¹) and chloramphenicol (35 μ g ml⁻¹). Subsequent cell culture and protein expression in minimal media, along with incorporation of selenomethionine, was incorporated during protein expression according to ref. 38.

Purification. Cell pellets from 3 l of cell culture were resuspended in 60 ml of PBS solution (140 mM NaCl, 2.5 mM KCl, 10 mM NaH₂PO₄, 2 mM KH₂PO₄) containing 10 mM imidazole (pH 8.0). Cells were lysed via sonication and centrifuged to remove cell debris. The supernatant was passed through a 0.2 μ m filter, and then incubated on a rocker platform at 4 °C for 1 h after adding 3 ml of resuspended nickel-NTA metal affinity resin (Invitrogen). After loading onto a gravity-fed column, the resin was washed with 45 ml of the same lysis buffer described above, and the protein was eluted from the column with three consecutive aliquots of PBS containing 150 mM imidazole (pH 8.0). Purified protein was concentrated to approximately 5–25 mg ml⁻¹ while buffer exchanging into 25 mM Tris (pH 7.5) and 200 mM NaCl and then further purified via size-exclusion chromatography using HiLoad 16/60 Superdex 200 column (GE).

Protein samples were then split in half; one sample was used directly for crystallization while the other had the His tag removed by an overnight digest with biotinylated thrombin (Novagen), before additional crystallization trials. The digested sample was incubated for 30 min with streptavidin-conjugated agarose (Novagen) to remove the thrombin. All samples were tested for purity and removal of the His tag via SDS-polyacrylamide gel electrophoresis. The final protein samples, both with and without the N-terminal poly-histidine affinity tag, were concentrated to values of 5–25 mg ml⁻¹ for crystallization trials.

Solution size and stability analysis. Proteins at a concentration of 4–10 mg ml⁻¹ were run over a Superdex 75 10/300 GL column (GE Healthcare) in 25 mM Tris pH 8.0 plus 100 or 750 mM NaCl at a rate of 0.4 ml min⁻¹ on an AKTApurime plus chromatography system (GE Healthcare). All fractions containing eluted toroid protein (visualized via electrophoretic gel analyses) were pooled, concentrated, and run over the column a second time to assess their solution oligomeric behaviour using protein with a minimal background of contaminants. Gel filtration standards (Bio-Rad) were run over the same column in matching buffer, and the ultraviolet trace of the proteins was overlaid onto the standards using UNICORN 5 software (GE Healthcare).

For measurements of protein stability using circular dichroism spectroscopy, purified recombinant toroid constructs were diluted to between 10 and 20 µM concentration and dialysed overnight into 10 mM potassium phosphate buffer at pH 8.0. Circular dichroism thermal denaturation experiments were performed on a JASCO J-815 circular dichroism spectrometer with a Peltier thermostat. Wavelength scans (190–250 nm) were performed for each construct at 20 °C and 95 °C. Additional thermal denaturation experiments were conducted by monitoring circular dichroism signal strength at 206 nm over a temperature range of 4–95 °C (0.1 cm path-length cell), with measurements taken every 2°. Sample temperature was allowed to equilibrate for 30 s before each measurement.

Crystallization and data collection. Purified proteins were initially tested for crystallization via sparse matrix screens in 96-well sitting drops using a mosquito (TTP LabTech). Crystallization conditions were then optimized with constructs that proved capable of crystallizing in larger 24-well hanging drops. Out of 11 constructs that were purified to homogeneity, 10 were crystallized, of which 5 yielded high quality X-ray diffraction that resulted in successful structure determination.

dTor_6x35L was crystallized in 160 mM sodium chloride, 100 mM Bis-Tris pH 8.5 and 24% (w/v) polyethylene glycol 3350 at a concentration of 26 mg ml⁻¹. The crystal was transferred to a solution containing 300 mM, then 500 mM sodium chloride and flash frozen in liquid nitrogen. Data were collected on a R-AXIS IV++ at wavelength 1.54 Å and processed on an HKL2000 (ref. 39).

dTor_6x35L(SeMet) was crystallized in 140 mM sodium chloride, 100 mM Tris pH 8.5 and 22% (w/v) polyethylene glycol 3350 at a concentration of 26 mg ml⁻¹. The crystal was transferred to a solution containing 300 mM, then 500 mM sodium chloride and flash frozen in liquid nitrogen. Data were collected at ALS Beamline 5.0.2 at wavelength 0.9794 Å and processed on an HKL2000 (ref. 39).

dTor_3x33L_2-2 was crystallized in two different conditions, producing two different crystal lattices. The first condition had 30% polyethylene glycol 3350, 100 mM Tris pH 6.5, 200 mM NaCl with a protein concentration of 1.8 mM. The protein was soaked in a 15% ethylene glycol cryoprotectant for 1 min before being flash frozen in liquid nitrogen. Data were collected on a Saturn 944+ (Rigaku) at wavelength 1.54 Å for 180° at $\varphi = 0$ and another 180° at $\varphi = 180$. Data were then processed on an HKL2000 (ref. 39) out to 1.85 Å in space group $P2_12_12_1$.

The second condition had 45% polyethylene glycol 400 and 100 mM Tris pH 7.7 with a protein concentration of 1.8 mM. Protein crystal was flash frozen without being cryoprotected. Data were collected on a Saturn 944+ (Rigaku) at wavelength 1.54 Å for 180° at $\varphi = 0$ and another 180° at $\varphi = 180$. Data were then processed on an HKL2000 (ref. 39) out to 1.85 Å in space group $P4_32_12$.

dTor_9x31L_sub was crystallized in 100 mM Tris pH 8.5 and 15% (v/v) ethanol at a concentration of 11.5 mg ml⁻¹. The crystal was transferred to a solution containing 75 mM Tris pH 8.5, 7.5% (v/v) ethanol and 25% (v/v) glycerol and flash frozen in liquid nitrogen. Data were collected at ALS Beamline 5.0.2 at wavelength 1.0 Å and processed on an HKL2000 (ref. 39) out to 2.9 Å in space group $P4_12_12/P4_32_12$.

dTor_9x31L was crystallized in 0.1 M sodium citrate pH 5.4 and 1.0 M ammonium phosphate monobasic at a concentration of 8.8 mg ml⁻¹ in 3 µl drops containing 1 µl protein and 2 µl well solution. The crystal was transferred to a solution containing the well plus 25% (v/v) glycerol and flash frozen in liquid nitrogen. Data were collected on a Saturn 944+ charge-coupled device at wavelength 1.54 Å and processed on an HKL2000 (ref. 39) out to 2.5 Å in space group $P2_12_12_1$.

dTor_12x31L was crystallized in 0.9 M sodium malonate pH 7.0, 0.1 M HEPES pH 7.0 and 0.5% Jeffamine ED-2001 pH 7.0 at a concentration of 8.8 mg ml⁻¹ in

2 µl drops containing 1 µl protein and 1 µl well solution. The crystal was transferred to a solution containing 0.675 M sodium malonate pH 7.0, 0.075 M HEPES pH 7.0, 0.375% Jeffamine ED-2001 pH 7.0 and 25% glycerol, and flash frozen in liquid nitrogen. Data were collected on a Saturn 944+ charge-coupled device at wavelength 1.54 Å and processed on an HKL2000 (ref. 39) out to 2.3 Å in space group R3:H.

Phasing and refinement. The dTor_6x35L and both dTor_3x33L_2-2 structures were solved by Molecular Replacement with Phaser⁴⁰ via CCP4i⁴¹ using the Rosetta-designed structure as a search model. The structures were then built and refined using Coot⁴² and Refmac5⁴³, respectively.

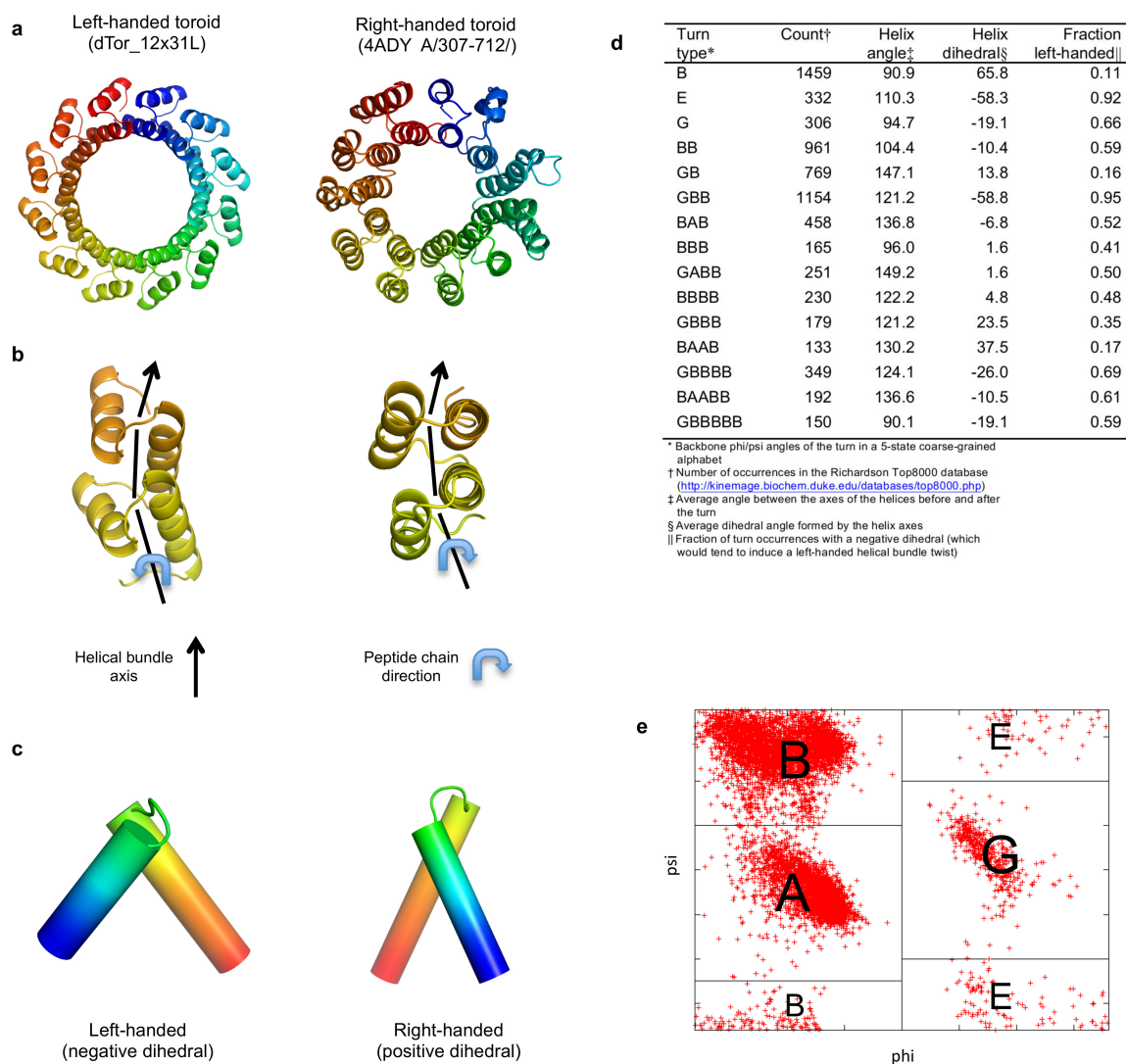
The structure of dTor_6x35L(SeMet) was solved by Molecular Replacement with Phaser⁴⁰ via PHENIX⁴⁴ using the best refined model of dTor_6x35L as a phasing model. The structure was then built and refined using Coot⁴² and PHENIX⁴⁵, respectively.

The structures of dTor_9x31L_sub and dTor_9x31L were solved by Molecular Replacement with Phaser⁴⁰ via PHENIX⁴⁴ using the Rosetta-designed structure as a search model. The structure was then built and refined using Coot⁴² and PHENIX⁴⁵, respectively.

The structure of dTor_12x31L was solved by Molecular Replacement with Phaser⁴⁰ via PHENIX⁴⁴ using a 4-repeat subunit the Rosetta-designed structure as a search model. The structure was then built and refined using Coot⁴² and PHENIX⁴⁵, respectively.

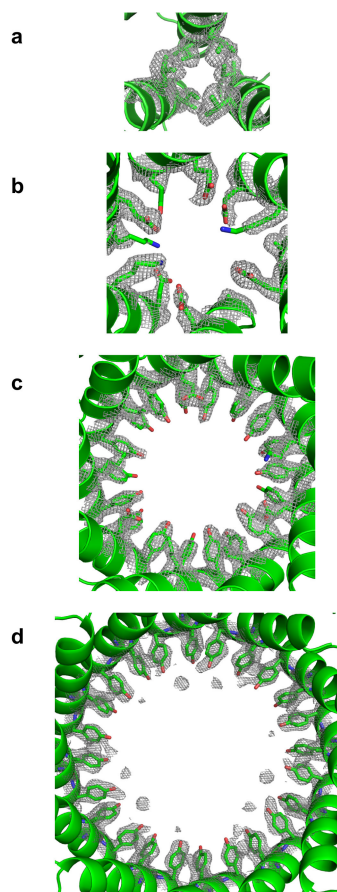
Final Ramachandran statistics after refinement were as follows (given as % preferred, % allowed, % outliers, respectively): dTor_6x35L(SeMet): 98.06, 1.94, 0.0; dTor_3x33L_2-2a: 99.48, 0.0, 0.52; dTor_3x33L_2-2b: 98.96, 0.52, 0.52; dTor_9x31L_sub: 98.31, 1.69, 0.0; dTor_9x31L: 99.28, 0.36, 0.36; dTor_12x31L: 99.0, 1.0, 0.0.

- Simons, K. T., Kooperberg, C., Huang, E. & Baker, D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.* **268**, 209–225 (1997).
- Dantas, G. *et al.* High-resolution structural and thermodynamic analysis of extreme stabilization of human procaryboxypeptidase by computational protein design. *J. Mol. Biol.* **366**, 1209–1221 (2007).
- Holm, L. & Sander, C. Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478–480 (1995).
- Sillitoe, I. *et al.* CATH: comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Res.* **43**, D376–D381 (2015).
- Fox, N. K., Brenner, S. E. & Chandonia, J. M. SCOPe: Structural Classification of Proteins—extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Res.* **42**, D304–D309 (2014).
- Cheng, H. *et al.* ECoD: an evolutionary classification of protein domains. *PLOS Comput. Biol.* **10**, e1003926 (2014).
- Mak, A. N., Lambert, A. R. & Stoddard, B. L. Folding, DNA recognition, and function of GIY-YIG endonucleases: crystal structures of R.Eco29kl. *Structure* **18**, 1321–1331 (2010).
- Walden, H. Selenium incorporation using recombinant techniques. *Acta Crystallogr. D* **66**, 352–357 (2010).
- Otwiński, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
- Skubák, P., Murshudov, G. N. & Pannu, N. S. Direct incorporation of experimental phase information in model refinement. *Acta Crystallogr. D* **60**, 2196–2201 (2004).
- Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
- Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352–367 (2012).
- He, J. *et al.* The structure of the 26S proteasome subunit Rpn2 reveals its PC repeat domain as a closed toroid of two concentric α -helical rings. *Structure* **20**, 513–521 (2012).

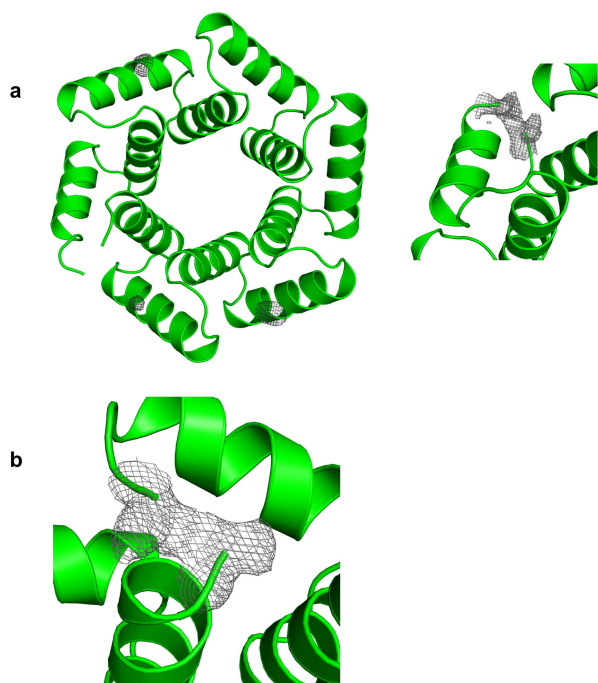


Extended Data Figure 1 | Handedness of α -helical bundles and helical linkers. **a**, Design dTor_12x31L, shown on the left, has a left-handed helical bundle. The native toroid on the right, which has a right-handed bundle, is taken from the Protein Data Bank structure 4ADY and corresponds to the PC repeat domain of the 26S proteasome subunit Rpn2 (ref. 46). **b**, The handedness of a helical bundle is determined by the twist direction of the polypeptide chain as it wraps around the axis of the helical bundle. **c**, Helical linkers characterized by a negative (positive) dihedral

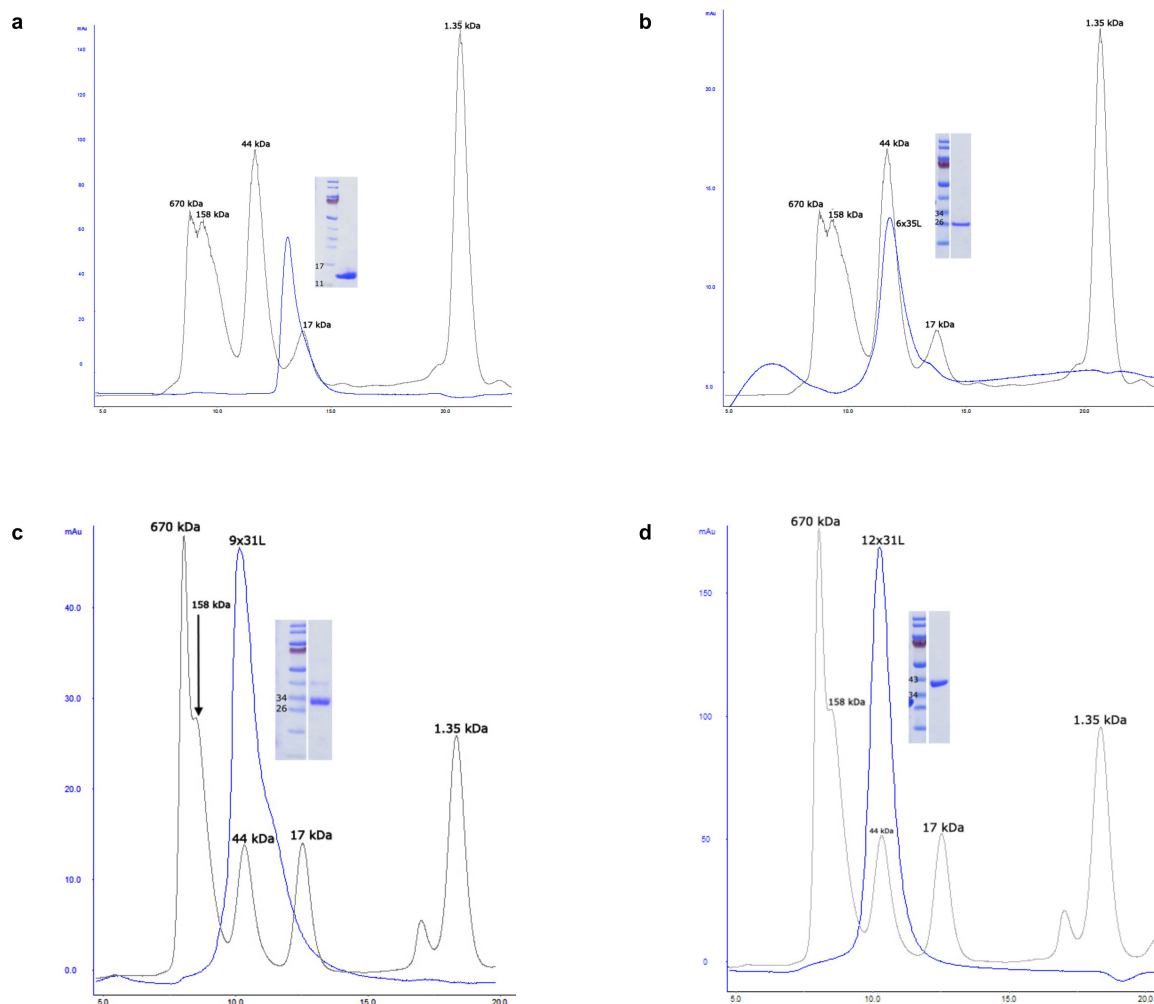
angle between the axes of the connected helices will, upon repetition, tend to impart a left-handed (right-handed) twist to the bundle. **d**, Geometrical properties of the most common short α -helical linkers in the structural database indicate that certain turn types (for example, 'E' and 'GBB') tend to form left-handed connections whereas others (for example, 'GB' and 'BAAB') are associated with right-handed connections. Turn types are classified by mapping their backbone torsion angles to a coarse-grained alphabet²⁷ as shown in **e**.



Extended Data Figure 2 | Unbiased $2F_o - F_c$ omit maps contoured around the side chains comprising the central pore regions for each crystallized toroid. The constructs shown are in the same order as in Fig. 3.

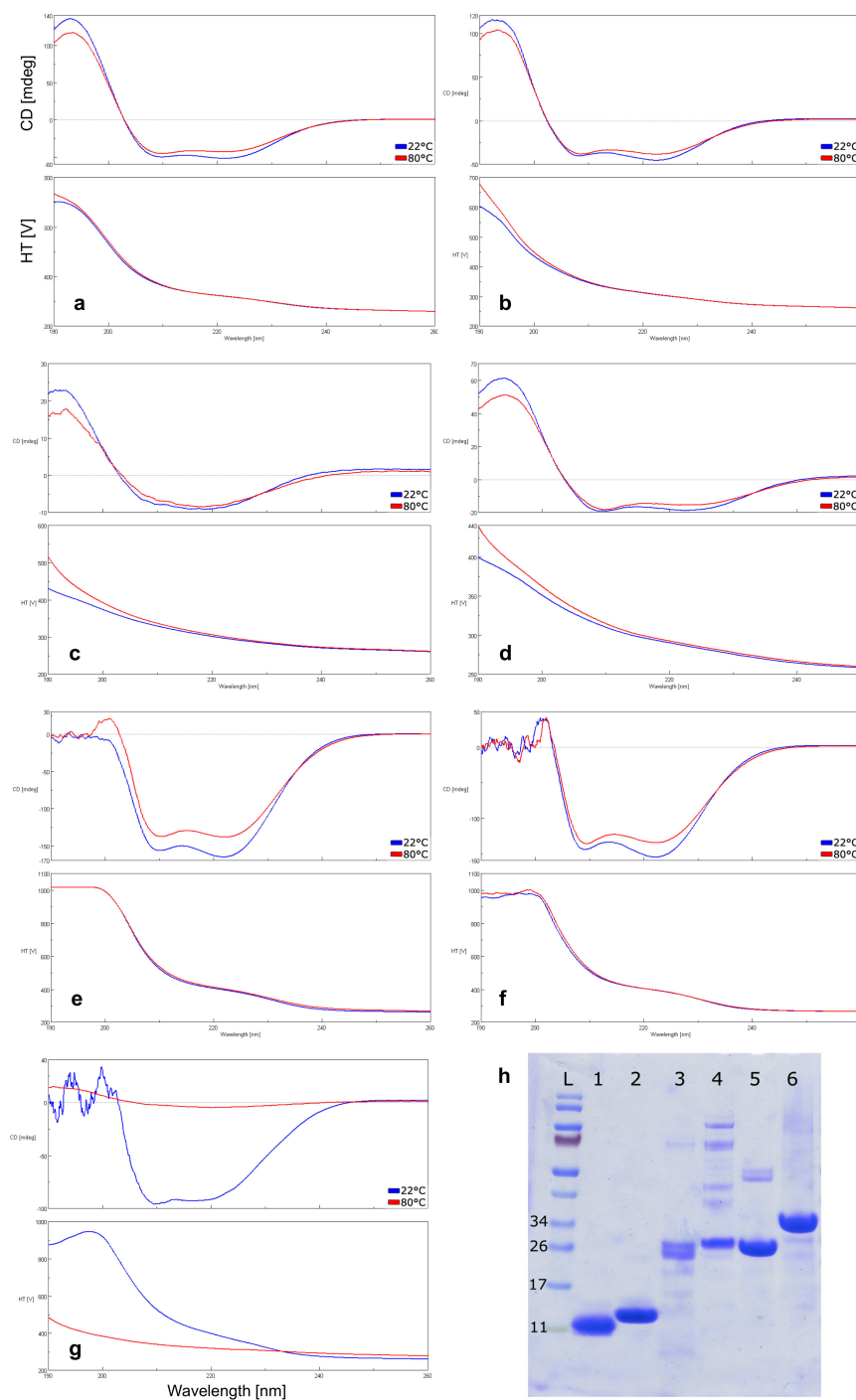


Extended Data Figure 3 | The crystallographic structures of highly symmetrical designed toroidal repeat proteins display rotational averaging in the crystal lattice. a, Electron difference density for construct dTor_6x35L. Left: anomalous difference Fourier peaks calculated from data collected from a crystal of selenomethionine-derivatized protein. Although only one methionine residue (at position 168) is present in the construct, strong anomalous difference peaks ($I/\sigma I$ greater than 4.0) are observed at equivalent positions within at least three modular repeats. Right: difference density extending across the modelled position of the N and C termini in the refined model, indicating partial occupancy at that position by a peptide bond. The other five equivalent positions around the toroidal protein structure display equivalent features of density, indicating that each position is occupied by a mixture of loops and protein termini. **b,** Electron density for construct dTor_12x31L, again calculated at a position corresponding to the refined N and C termini in the crystallographic model. As was observed for the hexameric toroid in **a**, the electron density indicates a mixture of loops and protein termini.



Extended Data Figure 4 | Size-exclusion chromatography elution profiles for the four designed toroids whose crystal structures were determined. The elution profiles (blue traces) shown correspond to runs in high (750 mM) NaCl for dTor_3x33L_2-2 (a) and dTor_6x35L (b), while the elution profiles for dTor_9x31L (c) and dTor_12x31L

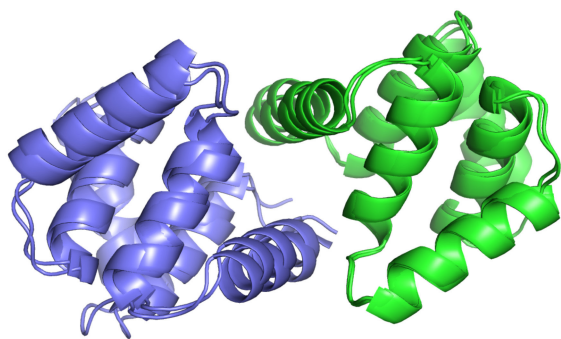
(d) correspond to runs in lower (150 mM) NaCl. The superimposed elution profiles of standard protein size markers (brown traces) correspond to runs at those same salt concentrations, conducted on the same column and day. The inset in each panel displays the migration and relative purity of each construct used for the analysis.



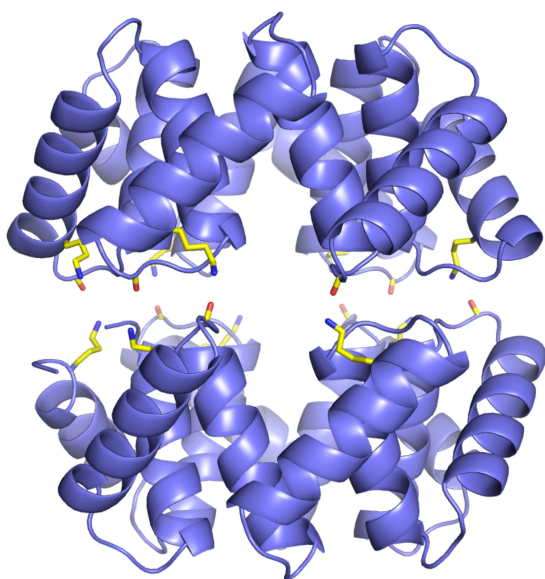
Extended Data Figure 5 | Purification and characterization of designed toroids. a–g, CD wavelength scan from 260 to 190 nm of several designed toroids and a positive control protein at 22 °C (blue) and 80 °C (red). a, dTor_9x31L_sub; b, dTor_3x33L_2-2; c, dTor_6x33R_1; d, dTor_6x35L; e, dTor_9x31L; f, dTor_12x31L; g, positive control. h, Bis-Tris gel (4–12%)

showing designed toroids immediately after metal affinity purification. Lane L, molecular mass protein standards (in kilodaltons); lane 1, dTor_9x31L_sub; lane 2, dTor_3x33L_2-2; lane 3, dTor_6x33R_1; lane 4, dTor_6x35L; lane 5, dTor_9x31L; lane 6, dTor_12x31L.

a



b



Extended Data Figure 6 | Potential dimerization interfaces observed in crystal packing interactions. **a**, Superposition of monomer–monomer packing interactions for the dTor_3x33L_2-2 design observed in two entirely different crystal forms. **b**, Stacking interactions between two dTor_6x35L subunits observed in the crystal structure; lysine residues interacting with backbone carbonyl groups in the partner monomer are shown in stick representation and coloured yellow along with their interaction partners.

Extended Data Table 1 | Characterization of designed constructs

ID	No. of repeats	Repeat length	Bundle handedness	Expressed*	Purified†	Oligomeric state‡	Crystals§	Structure
dTor_9x31L_sub¶	3	31	Left	Y	Y	M/D#	Y	Y
dTor_3x33L_1	3	33	Left	Y	Y		Y	N
dTor_3x33L_1-1	3	33	Left	Y	Y		N	
dTor_3x33L_2	3	33	Left	Y	Y		Y	N
dTor_3x33L_2-1	3	33	Left	Y	N			
dTor_3x33L_2-2	3	33	Left	Y	Y	D	Y	Y
dTor_3x33L_2-3	3	33	Left	Y	N			
dTor_3x33L_2-4	3	33	Left	Y	N			
dTor_3x33L_3	3	33	Left	Y	N/A			
dTor_6x33R_1	6	33	Right	Y	Y		Y	N
dTor_6x33R_1-1	6	33	Right	Y	N			
dTor_6x33R_1-2	6	33	Right	Y	N			
dTor_6x33R_1-3	6	33	Right	Y	N			
dTor_6x33R_2	6	33	Right	Y	N			
dTor_6x33R_3	6	33	Right	Y	N			
dTor_6x33R_4	6	33	Right	N				
dTor_6x35L	6	35	Left	Y	Y	D	Y	Y
dTor_6x35L(SeMet)	6	35	Left	Y	Y		Y	Y
dTor_9x31L	9	31	Left	Y	Y	M	Y	Y
dTor_12x31L	12	31	Left	Y	Y	M	Y	Y

*Construct was successfully overexpressed.

†Construct was successfully purified to homogeneity and concentrated to at least 1 mg ml⁻¹.

‡Dominant solution species, as assessed by size-exclusion chromatography (Extended Data Fig. 4); M, monomer; D, dimer.

§Construct crystallized.

||Crystals diffracted and structure determination was successful.

¶The 3-repeat subfragment of dTor_9x31L.

#Concentration-dependent monomer/dimer equilibrium.

Extended Data Table 2 | Crystallographic statistics

	dTor 6x35L	dTor 6x35L(SeMet)	dTor 3x33L 2-2a	dTor 3x33L 2-2b	dTor 9x31L sub	dTor 9x31L	dTor 12x31L
Data collection*							
Space group	C 2 2 21	C 2 2 21	P 21 21 21	P 43 21 2	P 43 21 2	P 21 21 21	C 2
Cell dimensions							
<i>a</i> , <i>b</i> , <i>c</i> (Å)	63.5, 85.3, 80.5	63.5, 85.1, 80.5	37.1, 68.6, 152.4	40.2, 40.2, 217.7	102.8, 102.8, 93.9	41.7, 72.0, 86.2	95.4, 119.4, 76.3
α , β , γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 110.9, 90.0
Resolution (Å)†	50.0-2.26 (2.30-2.26)	50.0-2.18 (2.26-2.18)	50.00-1.85 (1.90-1.85)	50-2.78 (2.88-2.78)	50.0-3.2 (3.3-3.2)	50.0-2.50 (2.54-2.50)	50.0-2.50 (2.54-2.50)
<i>R</i> _{merge}	0.045 (0.159)	0.059 (0.323)	0.056 (0.500)	0.048 (0.136)	0.056 (0.461)	0.079 (0.292)	0.048 (0.298)
<i>I</i> / <i>σ</i> <i>I</i>	39.9 (13.8)	29.7 (8.41)	20.3 (4.34)	27.0 (15.0)	31.3 (6.48)	30.4 (5.66)	27.2 (3.7)
Completeness (%)	98.1 (97.9)	99.7 (99.2)	90.6 (95.9)	98.9 (98.2)	100.0 (100.0)	99.2 (91.2)	98.9 (87.6)
Redundancy	3.8 (3.6)	13.7 (11.6)	6.0 (7.0)	12.3 (10.6)	14.8 (15.1)	10.0 (4.50)	3.7 (3.0)
Refinement							
Resolution (Å)		43.0-2.18 (2.23-2.18)	76.2-1.85 (1.90-1.85)	54.42-2.78 (2.85-2.78)	29.95-3.2 (3.7-3.2)	29.98-2.5 (2.6-2.5)	30.6-2.5 (2.59-2.50)
No. reflections		11137	29249	4760	8662	9355	27183
<i>R</i> _{work} / <i>R</i> _{free}		23.8/29.6	22.7/28.2	19.3/26.7	29.96/34.5	22.5/32.8	21.42/25.4
No. atoms							
Protein		1476	3038	1480	2292	2011	5608
Ligand/ion		-	8	-	-	-	-
Water		-	139	50	-	-	166
B-factors							
Protein		43.7	36.6	26	108.2	35.9	42.1
Ligand/ion		-	61	-	-	-	-
Water		-	52.4	56	-	-	43.8
R.m.s deviations							
Bond lengths (Å)		0.0142	0.017	0.017	0.002	0.008	0.002
Bond angles (°)		1.6908	1.708	1.918	0.5	1.038	0.49

*Each structure was determined from a single crystal.

†Highest resolution shell is shown in parenthesis.

CAREERS

POSTDOCS Independent funding brings freedom to move labs go.nature.com/gbeglg

ACADEMIA How to get a position as a staff scientist go.nature.com/qdpfc6

NATUREJOBS For the latest career listings and advice www.naturejobs.com



YENPITSU NEMOTO/GETTY

is quite normal for the reviewer comments to be published together with the original article. This also gives credit to the reviewer's ideas and contributes to the literature — it's an opinion from the expert and another perspective that could be very useful.

MAKE SOFTWARE ACCESSIBLE

Jean-Baptiste Mouret gave a six-limbed robot the ability to adapt quickly to a broken leg and other normally debilitating injuries by endowing it with the 'intuition' to try new approaches, such as hopping. The work, performed at the Pierre and Marie Curie University in Paris, graced *Nature's* cover in May. Mouret is now at the French Institute for Research in Computer Science and Automation near Nancy.

My dream would be to have an arXiv-like, free, centralized repository for source code. That way, it would be easy to reproduce and follow up on work that has been done.

Everything we do in science, including biology, physics and robotics, involves software. Now, when a paper is published, if we are lucky, there is a link to a web page somewhere with some version of the software. Most of the time there is nothing.

People often describe the algorithm in the paper, giving some equations and the main points of the software. But there is no way to check exactly how they integrated these equations or other details that don't fit into the paper. And many times, software has been written by a PhD student or a postdoc who has left the lab, and no one knows where a specific version of the software is. It is also very common to find papers for which the software has been available before, but has since disappeared in an update of a server somewhere.

Science would be much better if we had access to the software each time. Reviewers and journals should be asking for the source code. I don't think papers should be accepted without the software that corresponds to the analysis. It's like having a missing part of the paper. At the very least, it should be archived on the same web page as the paper or easily accessible from the paper itself.

But having one snapshot of the software in time is not enough. Software is a living thing. What we need is a central platform where we can submit bug fixes, improve the software and collaborate. This already happens for open-source projects. In computer science, we ►

INTERVIEWS

Big ideas for better science

We asked four researchers who made the news in 2015 what they would change about how science gets done.

REVEAL PEER REVIEWERS

As a biochemist at Seoul National University in South Korea, Jin-Soo Kim made headlines in 2015 for developing gene-editing methods that resulted in super-muscly pigs and new strains of tobacco, rice and lettuce.

Right now, peer review is usually blind in one direction. Reviewers know the authors' identities, but not the other way around. There is some merit to anonymity because reviewers can criticize a paper openly. But sometimes

the criticism is unfair.

Reviewers are sometimes competitors who may try to delay or block publication of a rival's work. They ask for more experiments, for additional data. Editors have to decide whether the comments are fair, but they cannot always make a proper judgement call. And in that case, a paper may be inappropriately delayed or rejected.

If the reviewer had to reveal his or her name after the paper was published, I think the reviews would be fairer. In other disciplines, such as the humanities and social sciences, it

► have very good platforms, such as GitHub, for developing software. Journals and institutions should partner with these companies. If we have a way to keep the software alive, it also makes it much easier to reproduce and continue the work.

This also implies that the software is open source, which I think is key for future science. Access to some software can cost €10,000 (US\$10,900) or more, which makes reproducing the research unattainable.

We have the technology to archive scientific software and link that software to papers. We just need the will.

BOOST WOMEN'S CAREERS

Planetary scientist Maria Cristina De Sanctis at the Institute for Space Astrophysics and Planetology in Rome was in charge of scanning the surface of the protoplanet Ceres using the orbiting Dawn spacecraft — the first time this asteroid-belt object has been examined up close.

I would change the way in which women are viewed in science — especially in the areas of technology development and instrumentation, because very few women are involved in those fields.

In Italy, sometimes school teachers and parents think that women and men belong in separate careers. For instance, secondary education includes classical schools based on the humanities and scientific schools based on the sciences and information technology. Most of the young women are in the classical institutes, whereas most young men are in the technical and scientific classes.

All of us should encourage girls to study sciences and support their education. This should start when parents are choosing toys, books and games — we should have the same approach for both boys and girls. Also, there should be some money reserved in grant programmes to support early-career women. I don't like the idea of having different programmes specifically for women — it can have unintended effects. But for particular fields, it could make sense in order to increase the proportion of women.

Women have a key role in the family. We need a more relaxed approach for considering things outside work. A woman who needs a few months to focus on something not related to work should be able to take that time off and then come back and refocus on her research.

In my experience and observations, women are generally less aggressive and may not seek to promote only themselves. This can be a real advantage in planetary science, where a large number of scientists come together for global collaborations and are not operating alone or in small groups. Having more women in higher positions could advance the science in better ways for the next generation.



Planetary scientist Maria Cristina De Sanctis.

TREAT SCIENTISTS AS HUMANS

Evolutionary biologist Danielle Edwards made the news in her home country of Australia when she turned down the prestigious Discovery Early Career Researcher Award, citing poor job prospects. Instead, Edwards, who specializes in herpetology, took a position as an assistant professor at the University of California, Merced.

I would change the way we gauge success in science from a quantitative approach to a more qualitative one. I think that would make science a safer place for people who have human needs. Time and time again, I've seen the shortcomings of the system play out in my life and in the lives of people who have decided to leave science.

We start out in a place where you have to work, work, work and your whole life is invested in your job. That really changes for some of us after we have children because we are forced to prioritize. Not having a safe place for those who value those non-work needs earlier on in their career results in less diversity in science. You get the drop out of women, the drop out of people who are first-generation college graduates, and the drop out of those from different backgrounds.

I don't think that working all the time equates to quality science. Some of the most productive researchers that I've ever met worked from 8 am until 5 pm, 5 days a week, and produced oodles of papers every year.

We need to change attitudes towards how we view success, the way we handle tenure, promotion and hiring, and the way we mentor students and postdocs. We need to recognize that scientists have basic needs for maintaining their family life, keeping healthy and not working long hours.

I say to my students, "Are you taking some time off?" I don't expect them to be in the lab late at night or on the weekends. I try to be as flexible and accepting of their human needs as I can be. A happy, healthy individual is going to produce quality work at the end of the day. It's a cost-benefit analysis: are you able to maintain that passion?

I come at this from multiple perspectives — I'm a first-generation college attendee, I grew up in a lower socio-economic area in Australia and I'm a woman in a relationship with a fellow scientist. I was told early in my career that as a woman, I was expected to work twice as hard. I know many colleagues whose trailing spouse, usually a woman, had to take a less-prestigious position than their partner, and their career was subsequently compromised. As a first-generation student, I've had people tell me that I didn't quite understand the academic life. And early on there was pressure from my family to stay close to home.

Sometimes that geographic pull is even stronger in people from different cultural backgrounds in which family is all important. That plays a huge part in siphoning out people from minority groups. We should be doing a better job in science to make sure people from different backgrounds are being encouraged. ■

Interviews by Kendall Powell. Interviews have been edited for length and clarity.

CORRECTION

The Careers feature 'Courage of conviction' (*Nature* **526**, 463–465; 2015) gave the wrong date for the conviction of Bradley Waldroup: the verdict was passed in 2009. The article also mischaracterized the part in the defence proceedings played by William Bernet. Bernet — together with James Walker — performed a complete psychiatric and neuropsychological profile of Waldroup and as a result identified that the defendant had a high-risk gene variant that, when coupled with his abusive childhood, could arguably increase his risk of violent behaviour. Bernet did not undertake any of the research linking this genetic variant to antisocial behaviour, as suggested by our article, but only presented a summary of extant scientific knowledge to the jury. Comments in the article also inadvertently could have been read as directly criticizing Bernet's testimony; this was not the intention and the text has now been corrected online to resolve this issue (see go.nature.com/xdi44d).

SINGLE LAYER I.T.

A collective celebration.

BY WILLIAM R. D. WOOD

The Waterford ball blazed into life and began its descent.
Sixty seconds.

The cheer and following swell of voices resonated in her temples and the sphenoid deep behind her nose. Shana was dizzy. The press of the crowd in Times Square was almost enough to hold the freezing New York night at bay. She closed the connection to the server back at the office and swiped the tablet to sleep.

Camera strobes fired from every direction. Family and long-time friends commemorated once-in-a-lifetime trips to the Big Apple. The hajj of the West. The air hung heavy with alcohol and musk, and she was reminded that many around her hadn't known each other a few hours ago. So many full of impure thoughts and desires. Celebrating and oblivious to those less fortunate.

Oblivious to those suffering pilgrims among them even now. Oblivious to her. Those less connected in a world population that had never been more connected.

Above, light rippled across the ball, overlaid patterns interspersed with bursts of colour. The promos had touted this year's theme for weeks, echoed tonight by every feed, private and commercial. The message was simple. *New Year. New York. New World.* A global culture that simultaneously embraced unity and individuality.

The promise of progress and understanding had never burned brighter, said all the newsfeeds. First North America, then the world.

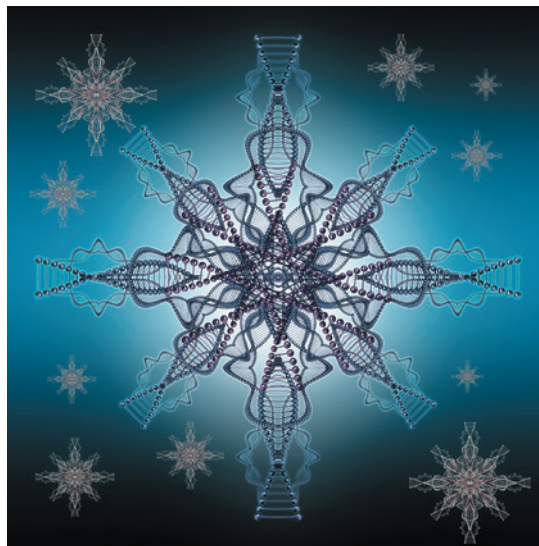
They were right, of course.

Fifty seconds.

Shana sneezed and wiped her nose. No one noticed. Ignorant of the contagion waiting patiently inside her. She was just another warm body, a drop of water in a churning sea. Her joints ached as the crowd swayed. Her lungs stung from the chill as she shared the breath of others and they shared hers. Her phone buzzed in her pocket. The familiar pattern of her fosters. They were proud but they would be prouder still.

Her eyes burned and her vision swam. She shuddered as the tiny vibration switched on inside her skull. Just like the test runs. The emitter went live right on time. Information and biology; blended, engineered.

A twenty-something frat-boy groped her as he and two buddies shouldered their way past. His drunken gaze met hers and he



shrugged, taking her picture and swiping it off to who knew where — his blog, a newsfeed, one of a billion sites.

Social networking. She smiled. Social *pluria* would finally be the promised but never truly delivered social *media*.

Forty seconds.

Winking at him just as he disappeared into the throng, she revelled in his split second of confusion. She'd not be sad to see an end to uninformed selfishness. His lack of connection to her and the countless people like her. The ignored. The alone. He'd know her and she, him.

People pulsed around her.

Displays as big as houses counted down from the towers, some giving the impression that their digits were raining down into the crowd. Scented vapours and smoke from something distinctly illegal tickled her nose and she sneezed. The thunder of voices surged and faded in waves. Her chest thrummed with each word as people joined together, counting in one voice. More and more in sync.

Every penny her fosters had spent sending her to school was worth more with each passing second.

Thirty.

Shana rubbed the bump behind her left ear where she'd administered the injection. She'd done the initial testing on herself. The engineered strains had done the rest. These

past few weeks she'd spent spreading the infection with her pamphlets. *Don't Miss New Year's Eve in*

Times Square! Grand Central, LaGuardia, the subways and every post office, park and public place within driving range. And she'd touched everything.

Twenty.

Lights throbbed, the voices around her a palpable force, threatening to buoy her from the ground. She coughed and dozens around her coughed as well, the reaction spreading outwards like a ripple in a lake of humanity.

Shana laughed. Experts had long remarked on the eerie structural similarity between a strand of DNA and a helical dipole antenna.

Ten.

It was as if nature had intended humanity to be a linked community from the very beginning. Higher order in the entropic background soup. People should be more than an hour, or a minute or a second in one another's day. Relationships were grander than an excuse to shove one's genetic material into another.

People staggered around her, struggling to process the sudden flood of information. That was to be expected. The disparity must be confusing.

The roar of voices faded.

Never again would a little girl watch her parents waste away without someone to hold her hand.

Five.

The crystal ball dropped its final inches as particles 3 million times smaller swarmed across blood-brain barriers. Mostly just NYC today. But this gift would spread, unstoppable. She'd been one of lucky ones but there were many more.

Countless more.

Four.

After a few days, no one would want to. Right now there were shouts for help echoed among those yet to be infected. Yet to be connected. But that would pass. Never again would a child depend on pedestrians for food.

Three.

A collective gasp rose from Times Square. Never again would anyone long for a single word.

Two.

Never again. Because. We. Are.

One. ■

ON NATURE.COM

Follow Futures:

@NatureFutures

go.nature.com/mtoodm

William R. D. Wood writes speculative fiction from a secret lair in the mountains of Virginia. You can find him online at www.williamrdwood.com.

ILLUSTRATION BY JACEY